Washington University in St. Louis
Washington University Open Scholarship

Arts & Sciences Electronic Theses and Dissertations

Arts & Sciences

Spring 5-15-2020

# Neural Mechanisms of Cognitive Individual Difference: An Investigation of the Human Connectome Project

Shelly Renee Cooper
*Washington University in St. Louis*

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

 Part of the Cognitive Psychology Commons, and the Neuroscience and Neurobiology Commons

## Recommended Citation

Cooper, Shelly Renee, "Neural Mechanisms of Cognitive Individual Difference: An Investigation of the Human Connectome Project" (2020). *Arts & Sciences Electronic Theses and Dissertations*. 2174.
https://openscholarship.wustl.edu/art_sci_etds/2174

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychological and Brain Sciences

Dissertation Examination Committee:
Todd Braver, Chair
Deanna Barch
Janine Bijsterbosch
Patrick Hill
Joshua Jackson

# Neural Mechanisms of Cognitive Individual Difference: An Investigation of the Human Connectome Project

by
Shelly R. Cooper

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2020
St. Louis, Missouri

© 2020, Shelly R. Cooper

# Table of Contents

List of Figures ..... iii
Acknowledgments ..... v
Abstract ..... vi
Chapter 1: Introduction ..... 1
Chapter 2: Methods ..... 11
2.1 Participants ..... 11
2.2 Neuroimaging Data and Tasks ..... 12
2.3 Network Assignment ..... 14
2.4 Behavioral Data ..... 19
2.5 Statistical Methods ..... 20
Chapter 3: Results ..... 34
3.1 Addressing Confounds Related to Family Structure of HCP Dataset ..... 34
3.2 Aim 1: 2x2 Analyses ..... 35
3.3 Aim 1: 4x5 Analyses ..... 36
3.4 Aim 2: 2x2 Analyses ..... 37
3.5 Aim 2: 4x5 Analyses ..... 49
Chapter 4: Discussion ..... 55
4.1 The Importance of Tasks ..... 59
4.2 The Quiet FPN and Negative CON ..... 61
4.3 Parcellation Methods and Levels of Analysis ..... 64
4.4 Latent Variable Models and Neuroimaging – The Good, The Bad, and The Future ..... 67
4.5 Implications for the study of Cognitive Control ..... 72
4.6 Conclusions ..... 75
References ..... 76
Appendix ..... 85

# List of Figures

Figure 1. Network Comparisons Per Parcellation Atlas ..... 16

Figure 2. Overlapping Networks..... 18

Figure 3. 2x2 Measurement Model Schematic ..... 27

Figure 4. 4x5 Measurement Model Schematic ..... 28

Figure 5. 2x2 Structural Models to List Sorting with Schaefer 100 ..... 40

Figure 6. 2x2 Structural Models to List Sorting with Schaefer 300 ..... 41

Figure 7. 2x2 Structural Models to List Sorting with Gordon Atlas ..... 42

Figure 8. 4x5 Structural Models to List Sorting with Schaefer 100 ..... 50

Supplement Figure 1 ..... 90

Supplement Figure 2 ..... 91

Supplement Figure 3 ..... 92

Supplement Figure 4 ..... 93

Supplement Figure 5 ..... 94

Supplement Figure 6 ..... 95

Supplement Figure 7 ..... 96

Supplement Figure 8 ..... 97

# List of Tables

Table 1. Number of Parcels Per Network ..... 22

Table 2. Parcels Adjusted for Ill-Scaling ..... 23

Table 3. Number of Parcels Per Model ..... 33

Table 4. Fit Indices of Aim 1 – 2x2 Analysis ..... 36

Table 5. Fit Indices of Aim 1 – 4x5 Analysis ..... 37

Table 6. Chi-Squared Difference Tests – 2x2 ..... 38

Table 7. Chi-Squared Difference Tests – 4x5 ..... 38

Table 8. Variances and Variances Explained – 2x2 ..... 45

Table 9. Variances and Variances Explained – 4x5 ..... 54

Supplement Table 1 ..... 85

Supplement Table 2 ..... 86

Supplement Table 3 ..... 87

Supplement Table 4 ..... 88

Supplement Table 5 ..... 89

# Acknowledgments

Data for this project were kindly made available via the Human Connectome Project with principal investigators Drs. David Van Essen and Kamil Ugurbil (grant number 1U54MH091657) funded by the NIH Blueprint for Neuroscience Research. Author SRC was also funded by the National Science Foundation Graduate Research Fellowship Program (DGE-1143954).

Shelly R. Cooper

*Washington University in St. Louis*

*May 2020*

ABSTRACT OF THE DISSERTATION

Neural Mechanisms of Cognitive Individual Difference: An Investigation of the Human Connectome Project by

Shelly R. Cooper

Doctor of Philosophy in Psychological and Brain Sciences

Washington University in St. Louis, 2020

Professor Todd Braver, Chair

Considering individual differences in task activation functional magnetic resonance imaging (t-fMRI) can be challenging because they may arise from variability in activity in brain regions, in the tasks themselves, or some combination thereof. Delineating sources of between-subjects variance is particularly important for cognitive control where task goals are at the forefront. Here we applied structural equation modeling (SEM) to the Human Connectome Project to examine if activity could be partitioned into separable brain and task individual difference dimensions. A series of SEMs were defined with varying numbers of latent factors, where the inputs were parcels of two cognitive control-related brain networks measured during two cognitive control-related task paradigms. Model comparisons favored the SEM where each network and task were specified separately. The same analyses were repeated with additional higher-order brain networks and tasks, and still the best-fitting model had latent factors for each task and network. Brain networks and task contexts are thus critical sources of individual differences, especially in the realm of cognitive control, and the t-fMRI signal can be decoupled accordingly. We further discuss the ramifications of considering different aspects of neuroimaging signals when interrogating brain-behavior relationships.

# Chapter 1: Introduction

A large component of cognitive neuroscience research has focused on the use of task-based functional magnetic resonance imaging (t-fMRI) as a tool to investigate the neural bases of various cognitive functions via tightly controlled experimental paradigms (e.g., is there a difference in mean neural activity between conditions or groups?). Yet important details get lost in this approach, simply due to within-group averaging across individuals. Consequently, translation of experimental findings into impactful therapeutics may ultimately fall short, especially in a domain like cognitive control, for which individual differences are thought to play a major role (Braver, Cole, & Yarkoni, 2010; Kane & Engle, 2002; Miyake et al., 2000). This discrepancy has recently led to large-scale efforts (e.g., NIMH Research Domain Criteria, or RDoC, initiative) dedicated to characterizing the spectrum of individual variation at multiple levels of granularity for various domains, including cognitive control. The goal of the current study is to validate and test the explanatory power of a highly applicable, but an under-utilized statistical methodology within neuroscience – structural equation modeling – to characterize individual differences in brain activation patterns and relate them to key issues in cognitive control.

Standard t-fMRI methods provide limited utility for characterizing the contribution of individual level variability in evoked fMRI brain activity patterns. One potential reason for this is that individual variability may be a characteristic of the brain network (or region) itself, yet present in a task-independent fashion. For instance, between-subjects variance patterns observed within a given network may persist across various task states. Additional, individual-level variation that is task-specific may *also* be present, but could be masked by task-independent

variation. Likewise, brain-behavior relationships may be preferentially observed if assessed during a particular task (state-like), or instead may be consistently present across multiple task contexts (trait-like). Accurate identification of brain-behavior relationships that operate in a more trait-like (i.e., stable, task-independent) versus those that are present in state-like manner will have important implications for understanding the continuum (or potential discontinuities) between healthy individual variation and neurocognitive impairments. It could also serve to increase validity in existing group-based comparisons through better control of individual-level variance.

The above issues are particularly salient for investigations of the neural mechanisms of cognitive control, a domain inherently dependent upon the task at hand. That is, cognitive control is defined by the ability to actively maintain particular task goals and update them accordingly. As a consequence, specific task demands are particularly relevant for cognitive control, which makes individual variation in cognitive control to be an especially likely candidate function that could exhibit state-like brain-behavior relationships (e.g., more task-related variance). A better understanding of the sources of individual variation that contribute to cognitive control function would have broad implications, as cognitive control is well-established to play a critical role in many task domains (e.g., attention, working memory, decision-making, reward processing, etc.; (Botvinick, 2007; Chiew & Braver, 2011; Gray, Chabris, & Braver, 2003; Kane, Bleckley, Conway, & Engle, 2001; Redick, 2014; Richmond, Redick, & Braver, 2015; Satterthwaite et al., 2007). Moreover, cognitive control is thought to be a central factor in a wide variety of mental health disorders and dysfunctions (e.g., schizophrenia, ADHD, Alzheimer's). Lastly, cognitive control has been clearly identified as a construct subject to substantial inter- and intra-individual differences in behaviorally focused investigations (Braver, 2012). As such, it has been identified

by the NIMH RDoC initiative as a target construct of interest. This is *not* to insinuate that there are *more* individual differences in cognitive control than in other domains, such as working memory, episodic memory, attention etc. Rather, cognitive control is a domain where inter-individual differences are thought to play a major role. Therefore, delineating dimensions that underlie individual differences in cognitive control is of interest not only from a basic science perspective, but also because of its clinical relevance. Further, lessons learned from this domain can then be applied to additional cognitive domains, enabling more direct comparisons across constructs. The purpose of the current study is to tease apart the between-subject variability of the t-fMRI BOLD signal into brain region-related and task-related dimensions, and to examine how these differentially correlate to behaviors both related and unrelated to cognitive control.

Prior investigations of individual differences in t-fMRI have been impeded by the analytical challenges associated with this endeavor. One limitation is that in much of the prior work, they have been assessed as an after-thought of a between-condition or between-group analysis, and via simple correlations (Pearson or Spearman; Yarkoni & Braver, 2010). Yet there are statistical frameworks optimized for the study of individual differences, mostly developed from within the field of psychometrics – of which, latent variable modeling methods such as Structural Equation Modeling or SEM, might be the most applicable. In SEM, observed (manifest) variables are linked to unobserved (latent) constructs via concurrent regression equations. This is done by comparing the variance-covariance matrix of an implied, researcher-specified model to the variance-covariance matrix of the observed data (Bollen, 1989; Kline, 2015). By mathematically modeling user-defined sources of between-subject variability, researchers can flexibly deploy a hypothesis-testing framework to simultaneously ask questions regarding: a) how individual variability across multiple manifest variables ought to organize into

latent individual differences dimensions (the measurement model), and b) how individual differences dimensions correlate to other latent dimensions and/or predict other observed behaviors (the structural model). Importantly, latent variables defined in SEM are considered "error-free" in that they reflect the variance shared by multiple manifest variables; they also enable shared variance to be "partialled out", if it can be attributed to other latent factors. As such, SEM procedures are especially adept at delineating and evaluating sources of individual differences, while simultaneously minimizing the influence of measurement error on the latent variables; thus, they lead to increased psychometric reliability. Likewise, since the latent constructs are theoretically specified and constrained, results are also thought to be more valid than traditional analyses. For more on using SEM on neuroimaging datasets, see Cooper, Jackson, Barch, & Braver (2019).

The advantages of SEM make it an ideal technique for the proposed characterization of individual difference dimensions in brain activation patterns. Specifically, it provides a flexible framework from which to partition individual variability in t-fMRI into latent constructs that separately reflect both brain networks and task contexts (as well as more global factors, such that these can be correlated with a range of individual differences dimensions (including but not limited to cognitive, psychosocial, and health-related outcomes). Therefore, the application of SEM to t-fMRI data has the potential to provide new insights regarding the degree to which the low end of functioning within a healthy population is continuous versus discontinuous with that observed in various clinical populations. Unfortunately, to date there has been very little integration of these individual difference-focused statistical methods with t-fMRI datasets because SEM requires large sample sizes (for fMRI) to be most validly deployed. A standard of n = 200 participants is often considered to be the minimum needed for SEM procedures

(Boomsma, 1985). Because typical neuroimaging studies are both labor and time intensive, acquiring such large datasets has previously been considered to be cost-prohibitive.

Yet it can be reasonably claimed that neuroimaging research is at the dawn of a new era. In particular, the recent large-scale, multi-center Human Connectome Project (HCP; https://www.humanconnectome.org/study/hcp-young-adult) yielded one of the very first datasets that enables a systematic and rigorous investigation of the neural mechanisms that underlie individual variation in human higher-cognitive functions (Van Essen et al., 2013). The HCP collected high quality, state-of-the-art neuroimaging data with comprehensive phenotyping (genetic, physiological, self-report, and behavioral information) on a demographically representative and genetically informed sample. Each subject participated in not only structural MRI, resting state fMRI, but also t-fMRI with a wide range of tasks, making it among the largest and richest publicly available datasets in existence. Since the HCP, other large-scale datasets have also been collected; yet, the HCP is particularly well-suited for an initial investigation into the utility of SEM approaches with regards to task fMRI. For instance, the UK Biobank is primarily focused on structural neuroimaging methods (https://www.ukbiobank.ac.uk/), with only a single short t-fMRI measure (Sudlow et al., 2015). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) project has a very wide age-range of participants (potentially increasing individual variation, but also making age a confounding factor); however, their t-fMRI procedures include only a single sensorimotor task plus movie watching, as opposed to multiple t-fMRI tasks (Shafto et al., 2014; http://www.cam-can.org/). Finally, while the currently ongoing Adolescent Brain Cognitive Development (ABCD; https://abcdstudy.org/) is following similar scanning procedures as the HCP, and will involve multiple t-fMRI measures collected on over 10,000 individuals in a longitudinal 10-year design, it has a primarily developmental focus

(Volkow et al., 2018). This adds in the additional complication of accounting for developmental differences and change effects; moreover, currently (at the time of this manuscript) only the first wave of data on children ages 9-10 is available.

The key question of the present project relates to the task contexts from which neuroimaging data is acquired. In order to address it properly, a dataset is required in which a large sample of participants are scanned while performing multiple task paradigms. Though usually not feasible (cost, time burden), the pooled "big data" resources from the HCP enabled each participant to be scanned during 7 different t-fMRI paradigms, two of which tap into cognitive control-related processes. Therefore, this unique HCP dataset is ideal for interrogating questions surrounding the neural circuitry that gives rise to individual differences, particularly as they relate to cognitive control.

An additional impediment to the adoption of latent variable model approaches in t-fMRI relates to the challenges in deciding between whole-brain voxel-wise and region-of-interest analyses. However, integrating recent developments from "network neuroscience" (Medaglia, Lynall, & Bassett, 2015; Sporns & Bassett, 2017) with individual differences research may help overcome this difficulty. A central insight that has emerged in the last decade is that brain regions are organized into functional networks, and that these networks show similar organization across both "resting" states and "task" states (Cole, Bassett, Power, Braver, & Petersen, 2014; Power, Schlaggar, & Petersen, 2014). Although the primary approach for defining networks has been on the basis of functional connectivity patterns, a critical assumption has been that these networks define an intrinsic level of organization of the brain, which should also be identifiable and useful for task activation studies. Newly developed parcellation algorithms yield a full set of cortical "nodes", postulated as unique, functionally meaningful sub-

units from which higher-level networks are defined (Gordon et al., 2016; Schaefer et al., 2018; Wig, Laumann, & Petersen, 2014). An innovative feature of the HCP dataset is that it has incorporated such parcellation schemes into an optional preprocessing pathway, making it easy to conduct analyses that utilize cortical parcels and functional networks as predefined building blocks.

Focusing on networks as the level of analysis seems like a particularly promising middle ground for examining individual differences in t-fMRI, as the preserved data in networks are more robust than typical voxel-wise analyses, yet are broader and potentially more functionally interpretable than region-of-interest analyses. Although certain networks have been strongly associated with cognitive control functions (at least at the group level), such as fronto-parietal (FPN) and cingulo-opercular networks (CON; Braver & Barch, 2006; Cole & Schneider, 2007; Dosenbach et al., 2007; 2006; Lerman-Sinkoff et al., 2017), there has not yet been a rigorous evaluation of the validity and functional utility of such brain networks for t-fMRI studies, particularly with respect to sensitivity to individual differences, both within specific networks and also across task contexts. The current project posits that the ability of brain networks to properly capture individual variation within and across tasks is an appropriate and powerful metric for such validation.

Although there is a rich history of t-fMRI studies examining smaller regions-of-interest that may be encompassed by these functional networks, to be clear, the current study is *not* assessing the claim that examining t-fMRI at the network level is better or worse than focusing on a particular node or region-of-interest. This is itself a very interesting question and worthy of investigation in future studies but is outside the scope of the current project.

The current study is divided into two specific aims. In the first aim, the goal is to test whether there are reliable individual differences that are brain network-specific, (i.e., related to key cognitive control networks, such as FPN and CON) and/or task-specific (i.e., related to key cognitive control paradigms, N-back and Relational Processing), utilizing SEM as the key analytic and inferential method. We hypothesize that partitioning the overall between-subject variability into more targeted nodes of individual difference (e.g., latent variables for each brain network and each task context) will provide a more internally consistent model of how BOLD data are inherently structured. That is, the best fitting model of t-fMRI BOLD data should be one that delineates task contexts and brain networks as separate sources of between-subject variability. Furthermore, we expect that the nature of t-fMRI BOLD data is such that even when expanding the focus to include a broader set of brain networks (e.g., dorsal attention network [DAN], default mode network [DMN]) and task contexts (e.g., Social Cognition, Language, and Gambling tasks), we will still find that the best fitting model is one that delineates the tasks and brain networks as separate sources of individual differences. Support of these hypotheses would facilitate the development of biologically constrained models of cognitive control. That is, future research may want to perform this type of variance decomposition procedure in order to create dimensions of cognitive control that are more faithful to the true internal structure of the individual differences contained in the t-fMRI signal. In turn, this can guide future hypothesis generation in a more targeted manner.

To be clear, the first aim of this study is focused entirely on the measurement model (e.g., how manifest variables organize into a latent factor structure), and the key data of interest are the overall model fit indices. Going forward, analyses only involving two brain networks and two cognitive tasks will be referred to as “2x2” whereas analyses involving the two additional brain

networks and three additional tasks will be referred to as “4x5”. As an aside, although the primary intention of the 2x2 analyses was to take a relatively narrow approach in targeting cognitive control, they also fulfilled a second goal of serving as a stepping-stone or proof-of-concept regarding the feasibility and utility of scaling-up to the larger 4x5 models. That is, if none of the 2x2 models converged, moving on to the 4x5 analyses would be exceedingly difficult.

The second aim extends the first by probing which of the neural activation latent variables reflecting individual difference dimensions (e.g., specific networks, specific tasks) best predict outcome measures that should be of theoretical relevance to cognitive control (for example, working memory). As with the first aim, the second aim has two subcomponents: 1) first, in the 2x2 setting which only includes a narrow set of cognitive control-related tasks (N-back and Relational Processing) and brain networks (FPN and CON) 2) then again with the 4x5 expanded set of tasks and brain networks. In this aim, there were three sets of outcome variables that vary in their supposed relationship to cognitive control. For the 2x2 phase, we hypothesized that all four latent constructs (two brain networks and two task contexts) would significantly predict outcome variables most strongly related to cognitive control, but exhibit a smaller effect sizes for the outcome variable expected to be only moderately related to cognitive control, and not significantly predict an outcome variable that should be unrelated to cognitive control. For the 4x5 expanded phase, we expect that the same relationships observed in the 2x2 will hold even in the presence of additional tasks and brain networks, although this is more exploratory in nature. In other words, the key focus of the second aim is to test whether brain networks and task contexts are both important dimensions of individual differences in cognitive control in terms of predicting relevant outcome variables, above and beyond other classically higher-order brain

networks and other general cognitive tasks. The primary focus of this aim will be to carefully examine particular parameter estimates/regression coefficients across various models in order to evaluate if separating the sources of individual differences results in any gains (or losses) in explanatory power. Knowledge of this nature is essential for precision medicine efforts, as support for this hypothesis would indicate that future interventions targeting neurocognitive impairment might only be effective in specific environmental contexts.

# Chapter 2: Methods

The primary approach of this manuscript is to apply the modern latent variable framework from the psychometric literature to neuroimaging data in order to better characterize the neural factors that underlie individual differences in cognitive control. The following sections describe the participants, neuroimaging tasks and data processing, then providing greater detail about the statistical methodology. Note that in order to facilitate open access to all aspects of the research lifecycle, most activities related to this project (preprocessing scripts, analysis scripts, publications etc.) are contained on Open Science Framework (https://osf.io/a6x5b/), and all preprocessed neuroimaging and behavioral data is publicly available through the HCP website (https://www.humanconnectome.org/).

## 2.1 Participants

The HCP Healthy Young Adult full release dataset (n = 1200) was used for all aspects of this project, and included healthy participants ranging from 22-35 years of age. Although a broad set of imaging and other data were collected for the HCP, the current project focuses on t-fMRI and associated behavioral outcomes. As such, participants were included if they had neuroimaging data available for each of the 5 cognitive tasks and completed the three out-of-scanner tasks (described below), resulting in a final sample size of n = 1005. Note that family structure was not taken into account for the primary analyses; however, supplemental analyses described below tested the validity of this approach to inference. Here the HCP is considered an archival dataset, and no new participants were recruited for this project. For more details regarding HCP participant recruitment and informed consent processes, please see Van Essen et al., (2013).

## 2.2 Neuroimaging Data and Tasks

Detailed aspects of the neuroimaging data acquisition and preprocessing protocol are available both on the HCP website (www.humanconnectome.org) and in various publications (Barch et al., 2013; Glasser et al., 2013; Ugurbil et al., 2013). Broadly however, HCP data were collected on a Siemens 3T Skyra and acquisition parameters feature whole-brain coverage, a 32-channel head coil, multi-band acceleration, and high spatial and temporal resolution (2 mm voxels, <1s TR).

The HCP protocol included 7 t-fMRI paradigms, but two were excluded from the current project: Motor and Emotion tasks. The Motor task was excluded because it exhibits minimal between-subjects variability in the corresponding out-of-scanner motor behavioral measures. The Emotion task was excluded because its utility was primarily for engaging subcortical limbic regions, especially amygdala (Hariri, Tessitore, Mattay, Fera, & Weinberger, 2002). Currently, the available parcellation algorithms are thus far best suited for cortical networks making it challenging to know how to incorporate subcortical regions into the relevant brain networks. Of note, this is an active area of research and future parcellation algorithms may soon be able to account for the subcortex and cerebellum (Seitzman et al., 2018). Consequently for the current study, the emphasis was on the five remaining task paradigms: N-back, Relational Processing, Gambling, Language, and Social Cognition. Comprehensive rationales for HCP task selection, task descriptions, and all relevant task parameters have been extensively reported in Barch et al., (2013). Below, brief descriptions of the tasks are provided describing the key aspects and activation contrasts from each of the five task paradigms.

*N-back:* The N-back is a well-established working memory (WM) paradigm, which includes blocked 2-back (high WM-load) and 0-back (low WM-load) conditions, performed with

a variety of stimulus types that varied across blocks. The current study focuses on activation that should isolate WM load effects, via the 2-back - 0-back contrast (cope 11), collapsing across stimulus type. This task is included in both the 2x2 and 4x5 set of analyses.

*Relational Processing*: This task engages higher-cognitive processes used in analogical reasoning, such as integration within WM. In the relational blocks, the dimension along which one pair of objects differs (e.g., texture) must be extracted (and maintained in WM), and then compared with another pair of objects to determine if the latter vary on the same or different dimension. In match blocks, the judgment is just whether a bottom object matches either of the top objects on the specified dimension (shape or texture). The current study utilizes the activation present in the relational - match contrast (cope 4) to isolate relational processing effects. This task is included in both the 2x2 and 4x5 set of analyses.

*Gambling*: This task involves guessing card numerical values, with monetary rewards and punishments provided as feedback, in blocked mostly-reward and mostly-punishment conditions (Delgado, Nystrom, Fissell, Noll, & Fiez, 2000). To focus on these differential reward effects, the current study focuses on activation present in the reward - punishment contrast (cope 6). This task and data are only included the 4x5 set of analyses.

*Language*: This task requires participants to process auditorily-presented and semantically challenging stories in order to answer later comprehension questions, with story task blocks alternating with math blocks of matched length and difficulty (followed by comprehension questions; task adapted from (Binder et al., 2011). To focus on these differential language-related effects, the current study focuses on activation present in the story - math contrast (cope 4). This task and data are only included the 4x5 set of analyses.

*Social Cognition*: This task involves presentation of short videos depicting geometric shapes moving in ways that appear to express either social interactions (i.e., inferring intentionality, sometimes referred to as involving Theory of Mind; TOM) or random trajectories, with participants making a judgment regarding which type of pattern occurred (video clips adapted from Castelli, Happé, Frith, & Frith, 2000; Wheatley, Milleville, & Martin, 2007). To isolate these social interaction processes, the current study focuses on activation present in the social (or TOM) - random contrast (cope 6). This task and data are only included the 4x5 set of analyses.

## 2.3 Network Assignment

A key aspect of the proposed methodology is to treat functional networks (rather than voxels or regions-of-interest) as the primary unit of analysis, enabling significant data reduction while concurrently evaluating the validity of this approach. Each network is composed of a set of cortical parcels (treated as "nodes" of the network) defined from a parcellation algorithm. In general, these parcellations take coordinates delineating boundaries of individual parcels and apply them to individual subject t-fMRI data as a mask, thus individual parcels reflect the average BOLD signal across the set of voxels comprising the parcel. Each parcel is assigned as belonging to a network. Activation parameter estimates (in terms of percent signal change, defined from the HCP preprocessing pipeline) are then provided for each parcel, in each task contrast, for each participant.

There are now several different methods for defining these coordinate boundaries (and thus different parcellation algorithms; Glasser et al., 2016; Gordon et al., 2016; Power et al., 2011; Schaefer et al., 2018). Interested readers can find relevant information (such as parcel coordinates, labels etc.) and code for each of these parcellations at the following locations:

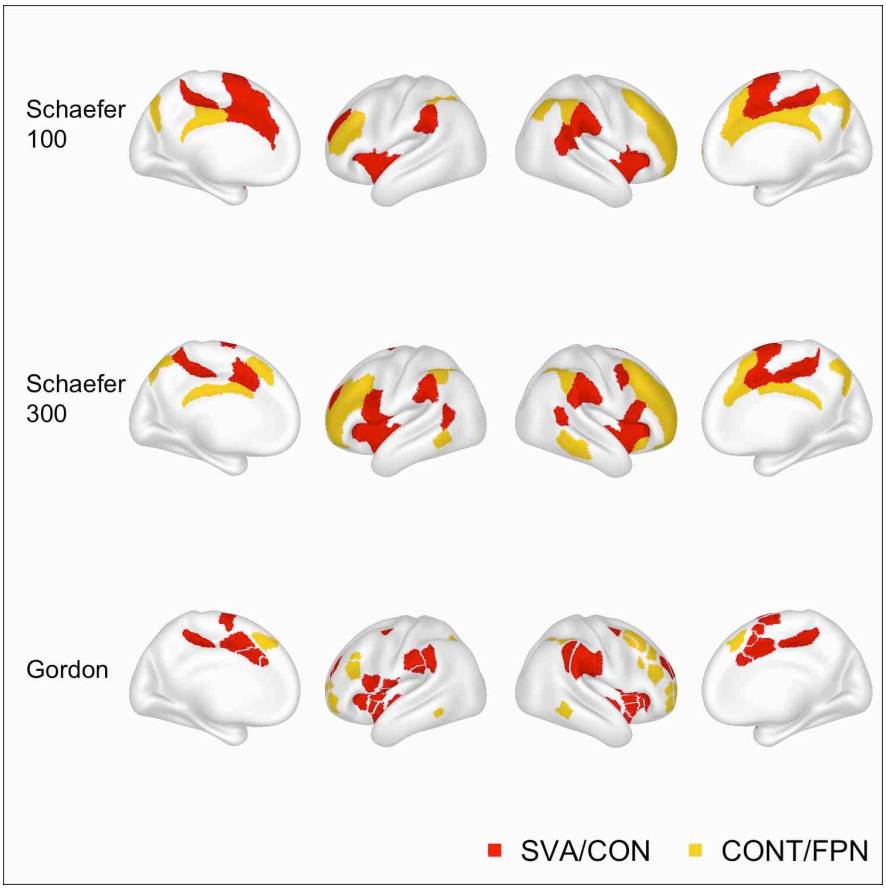
https://sites.wustl.edu/petersenschlaggarlab/resources/ for Power et al. (2011) and Gordon et al. (2016); supplementary information (online version of manuscript only) for Glasser et al. (2016); and https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal for Schaefer et al. (2018). If these parcellation mechanisms are indeed tapping the same underlying networks, then the inferences one might make from an analysis with one parcellation scheme should mirror the inferences one would make if replicating that analysis using a different parcellation scheme. To test this, the 2x2 analyses (aims 1 and 2) were performed using the Gordon et al. (2016) and the Schaefer et al. (2018) parcellations (note going forward these will be referred to as "Gordon/Schaefer parcellation", "Gordon/Schaefer parcels", or "Gordon/Schaefer atlas"). Additionally, the Schaefer parcellation has the option of specifying how many parcels should be defined. The current project uses the Schaefer 300 parcels in order to roughly match the number of Gordon parcels ($n_{GordonParcels} = 333$), as well as the Schaefer 100 parcels. The Schaefer 100 was chosen because it adds an element of extra data reduction. Again however, we expect results to be concordant across the three parcellation methods.

In all, four functional networks were examined in this study. Below, brief descriptions of the four networks are provided describing basic anatomical components and functional relevance. Italicized labels reflect the Schaefer atlas labeling.

*Control Network (Cont):* The Cont network anatomically maps to lateral prefrontal and frontal cortices and lateral posterior parietal cortex, including the intraparietal sulcus. In the Gordon atlas, as well is in a large portion of the literature, this is referred to as the frontoparietal network (FPN; however, we will keep with the Schaefer labeling for the duration of this article). This network has been extensively linked to cognitive control processes, especially showing

error-related activation and start-of-task engagement (Dosenbach et al., 2006; 2007; Gratton et al., 2016), with some even considering it a "flexible hub" of control (Cole et al., 2013). For further reading on this network, see Marek & Dosenbach (2018). We therefore consider this network to be one of the cognitive control networks in the current study, and it is used in both the 2x2 and 4x5 analyses. In Schaefer 100 there are 13 Cont parcels; in Schaefer 300 there are 40 Cont parcels; and in Gordon there are 24 FPN parcels. Figure 1 illustrates this network across the three parcellations.

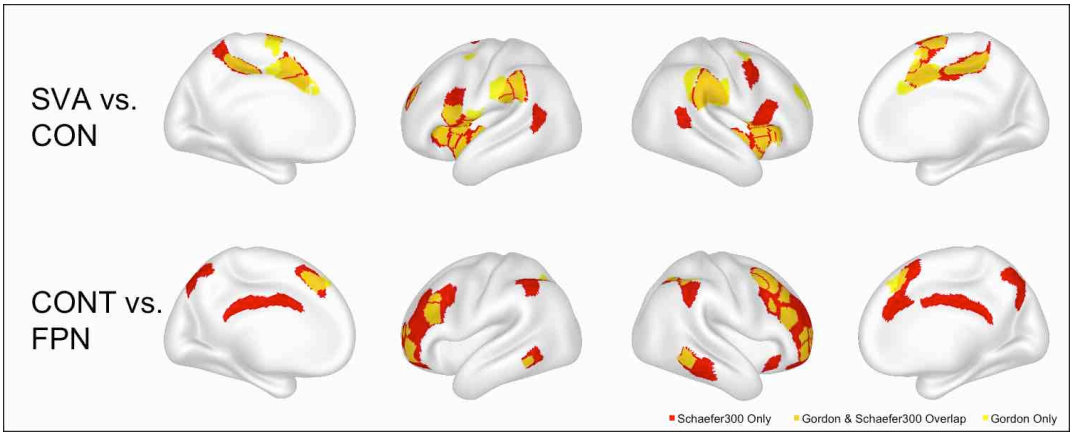
**Figure 1. Network Comparisons Per Parcellation Atlas**




*Salience Ventral Attention Network (SalVenAttn or SVA):* This network is comprised of regions in the dorsal anterior cingulate, as well as the anterior insula and frontal operculum. The

labeling of this network is particularly confusing, however. The Schaefer atlases are an extension of Yeo et al. (2011), who label their network as the "ventral attention network" and note that this corresponds with Fox, Corbetta, Snyder, Vincent, & Raichle (2006). Yet as Yeo and colleagues concede, what they call the "ventral attention network" is an amalgam of sorts of the cingulo-operuclar (CON) network and Salience network. Sometimes, the literature refers to a Salience network, but the anatomical correlates very closely mirror the CON (for example, see Seeley et al., 2007). Others consider the CON and Salience to be separable networks. In fact, the Gordon atlas does include a separate Salience network, however it only contains 4 parcels, compared to their 40 CON parcels. To maintain simplicity, we consider the SVA to be roughly analogous to the CON in the Gordon atlas. In Schaefer 100 there are 12 SVA parcels; in Schaefer 300 there are 34 SVA parcels; and in Gordon there are 40 CON parcels. Like the Cont (FPN), this network has been expressly related to cognitive control. In contrast to the Cont (FPN), however, the SVA (CON/Salience) has been shown to engage in a more sustained fashion suggesting it contributes to cognitive control via tonic alertness (Dosenbach et al., 2006; 2007; Sadaghiani & D'Esposito, 2015). We thus consider the SVA to be the second cognitive control network in the current study, and it is used in both the 2x2 and 4x5 sets of analyses. See Figure 1 for how this network appears across the different atlases. Figure 2 shows the degree of closeness in overlapping networks, specifically for the Schaefer 300 and Gordon atlases since they have similar numbers of parcels (300 and 333, respectively).

**Figure 2. Overlapping Networks**




*Dorsal Attention Network (DAN):* The DAN includes the bilateral intraparietal sulcus and frontal eye fields (Fox et al., 2006) and is primarily concerned with visuospatial attention, especially in regards to using a top-down cue to bias attention (Corbetta & Shulman, 2002). Much of the literature involving the DAN has been principally related to selective attention, rather than cognitive control, *per se*. As such, the current study considers the DAN to be a higher-order cognitive network, but not explicitly a cognitive control network. It is only examined in the 4x5 analyses with the Schaefer 100 atlas and consists of 15 parcels.

*Default Mode Network (DMN):* The DMN includes the posterior cingulate cortex, ventral anterior cingulate cortex, and medial prefrontal cortex (bilaterally). It is unique in that increased activation in the DMN occurs at rest, whereas it is "deactivated" or not as strongly engaged during goal-directed behavior (Greicius, Krasnow, Reiss, & Menon, 2003; Raichle et al., 2001). Here, the DMN is included in the 4x5 set of analyses as an interesting control network such that we expect a negative relationship between the DMN and a given outcome. In the Schaefer 100 atlas, the DMN has 24 parcels.

## 2.4 Behavioral Data

The behavioral outcome measures used in aim 2 were selected based on availability in the HCP dataset and theoretical relevance. Amongst the plethora of behavioral outcomes to choose from in the HCP dataset, three out-of-scanner measures were selected due to their varying degrees of theoretical relevance to cognitive control. Note that in-scanner task-associated behaviors were not considered. Behavioral performance on in-scanner tasks would be expected to be related to participant "states" (e.g., fatigue, mood, arousal) and traits, and may directly reflect some activation patterns (e.g., individuals making more errors might show stronger error-related patterns in SVA/CON networks; Yarkoni & Braver, 2010). This can sometimes lead to accidental statistical double dipping, and thus in-scanner performances were not taken into account in this study. We chose working memory (WM) to be the domain of most relevance to cognitive control (Kane et al., 2001; Redick, 2014; Richmond et al., 2015). Therefore, we expect any individual differences captured by cognitive control-related task states and brain networks to strongly predict WM. The current study uses the NIH Toolbox List Sorting Task (age-adjusted; Tulsky et al., 2014) as the WM measure.

As the List Sorting WM measure ultimately tests convergent validity, we then chose two additional constructs which were hypothesized to have varying levels of discriminant validity. First, we chose the Openness dimension from the NEO-FFI (McCrae & Costa, 2004), as Openness has been shown to positively correlate with IQ at around .4 (Goff & Ackerman, 1992), and it has been theorized that cognitive control is related to intelligence, especially fluid intelligence (gF; Duncan, Emslie, Williams, Johnson, & Freer, 1996; Gray et al., 2003; Kane & Engle, 2002). Further, IQ and WM have been shown to be related, but independent constructs (correlation of .48; Ackerman, Beier, & Boyle, 2005). We therefore expected that there could be

some moderate relationships between cognitive control-related networks and tasks and Openness, but also expected they would be weaker since Openness comes from the personality domain rather than being an index of cognitive ability (as WM & gF). To contrast, we include an additional 2x2 analysis where Openness is instead replaced with the Penn Progressive Matrices (PMAT; Bilker et al., 2012), which taps into gF in a more direct cognitive ability manner. Yet to reiterate, our aim was to find a construct where cognitive control-related networks and tasks would demonstrate smaller effect sizes, and thus prioritized the Openness dimension. As such, analyses with the PMAT are limited to the 2x2 with only the Schaefer 100 atlas, as inclusion in the full suite of analyses is beyond the scope of the current study.

Lastly, we selected a measure from the Motor domain of the NIH Toolbox – Grip Strength – as the third primary outcome measure (Reuben et al., 2013). In this task, participants squeeze a dynamometer to obtain a measure of grip strength force. Though Grip Strength has been shown to be related to some elements of cognitive functioning, these studies tend to focus on aging populations (Viscogliosi, Di Bernardo, Ettorre, & Chiriac, 2017). For example, a recent study from the UK Biobank sample showed a relationship between Grip Strength and memory and reasoning, but the mean age of the healthy sample was 56.49 (Firth et al., 2018). Since the HCP cohort is quite a bit younger than this sample (Van Essen et al., 2013), we expected Grip Strength to thus show the most divergent validity with regards to cognitive control brain networks and task states.

## 2.5 Statistical Methods

The current project uses a series of latent variable models, SEMs in particular, to test if there are reliable network-specific individual differences in the full HCP dataset, and if they persist across task states. For each set of analyses (2x2 per parcellation method and 4x5 with

Schaefer 100 parcels), a total of four models were defined, wherein each of the four researcher-specified models reflects a particular hypothesis about the organization of the underlying between-subject variability of t-fMRI BOLD data. The input for all SEMs was the same: t-fMRI parcels from each network in each of the task contexts (see Table 1 for number of parcels per network).

**Table 1. Number of Parcels Per Network**

| Atlas | Network | Number of Parcels | Cognitive Control Network? (Y/N) |
|---|---|---|---|
| Schaefer 100 | Salience Ventral Attention Network (SVA) | 12 | Yes |
| | Control Network (Cont) | 13 | Yes |
| | Dorsal Attention Network (DAN) | 15 | No |
| | Default Mode Network (DMN) | 24 | No |
| *Total Cognitive Control* | | *26* | |
| *Total All* | | *64* | |
| Schaefer 300 | Salience Ventral Attention Network (SVA) | 34 | Yes |
| | Control Network (Cont) | 40 | Yes |
| *Total* | | *74* | |
| Gordon | CinguloOpercular Network (CON) | 40 | Yes |
| | FrontoParietal Network (FPN) | 24 | Yes |
| *Total* | | *64* | |

However, before using the parcel data, a brief data cleaning procedure was done to correct for the "ill scaling problem". SEM relies on variance-covariance matrices. As such, large discrepancies in variances amongst manifest variables (here, parcels) can be problematic for any iterative estimation process, such as the maximum likelihood estimation used here. Thus, a common practice is to correct for this by multiplying or dividing by a scalar in order to improve the properties of the variance-covariance matrix, and ideally, they should be within a factor of 10 with each other. This was done for each of the four datasets used here (Gordon 2x2, Schaefer

300 2x2, Schaefer 100 2x2, and Schaefer 100 4x5), and Table 2 identifies which parcels were adjusted.

**Table 2. Parcels Adjusted for Ill-Scaling**

| Atlas | Network | Task | Parcel ID |
|---|---|---|---|
| 2x2 Analysis | | | |
| Schaefer 100 | *NA* | *NA* | *None* |
| Schaefer300 | CONT | N-back | LH_Cont_PFCl_5_nbk |
| Gordon | CON | N-back | L_CinguloOperc_ID147_nbk |
| | CON | N-back | L_CinguloOperc_ID28_nbk |
| | FPN | N-back | L_FrontoParietal_ID108_nbk |
| | FPN | N-back | L_FrontoParietal_ID109_nbk |
| | FPN | N-back | L_FrontoParietal_ID149_nbk |
| | FPN | N-back | L_FrontoParietal_ID7_nbk |
| | CON | Relational | L_CinguloOperc_ID28_rel |
| | FPN | Relational | L_FrontoParietal_ID149_rel |
| 4x5 Analysis | | | |
| Schaefer 100 | CONT | Social Cognition | LH_Cont_PFCl_1_socialcog |
| | DMN | Language | LH_Default_PCC_1_language |
| | DMN | Language | RH_Default_PCC_1_language |

*CONT – Control network; CON – Cingulo-Opercular Network; FPN – Fronto-Parietal Network; DMN – Default Mode Network.*

Below describes how each of the four measurement models were defined (see Figures 3 and 4 for schematic path diagrams of these four competing models for the 2x2 and 4x5 analyses, respectively), the hypothesis tested by the model, and the implications should the model be considered the "best fitting model" compared to the remaining three (additionally, see Table 3 for number of parcels per model):

*"Full Model" or "Full Bifactor Model"*: The full model was defined such that each task and each brain network had their own dedicated latent factors. That is, four latent variables were

specified in the 2x2 analysis: SVA/CON, Cont/FPN, N-back, and Relational Processing. In the 4x5, nine latent variables were specified: SVA/CON, Cont/FPN, DMN, DAN, N-back, Relational Processing, Social Cognition, Language, and Gambling. Importantly, the correlation amongst all latent variables was fixed to zero, and each parcel was allowed to simultaneously load onto two latent factors – one relating to the appropriate task and one relating to the appropriate brain network. This setup is known as a "bifactor SEM" and ensured that the between-subject variance of a single parcel was partitioned (or partialled) into brain networks/task contexts appropriately. For a schematic representation of this model, see panel D in Figures 3 and 4.

The full bifactor model reflects the hypothesis that both brain networks and task contexts are important dimensions of cognitive individual difference. In this full model, a network-specific latent variable, say the SVA/CON, is interpreted as the between-subject variance shared by all parcels in the SVA/CON (within network), after controlling for variance due to the different task states (i.e., isolating the task-independent variance). Conversely, a task-specific latent variable, say the N-back, captures between-subject variance shared across all cortical parcels measured in the N-back (within task), after removing variance due to each specific different brain network (i.e., isolating brain network-independent variance). If both brain networks and task contexts are separate sources of individual differences, then this full model should yield the best fit compared to the other three. While it may seem almost intuitive that this should be the case, given that the nature of t-fMRI is to induce particular task-related behaviors to better understand the neural mechanisms underlying these behaviors, to our knowledge this has not been formally studied or validated.

*"Null Model"*: An alternative to the full model described above is a far more parsimonious account where every parcel loads onto a single global latent variable (not a bifactor model). This null model reflects the hypothesis that while there may be shared variance across all parcels from all networks and tasks, neither brain networks nor task contexts are separable, independent dimensions of cognitive individual difference. This global factor can be thought of in the same manner as the first component of a principal component analysis. A single global latent factor would not be particularly informative since it would be difficult to determine the source of the shared variability. If this null model were the best fitting model, it would instead suggest that brain activation patterns are so intertwined with task-imposed variance, that they are not able to be decoupled. This model corresponds to panel A in Figures 3 and 4.

*"Partial Brain Model"*: Here, a bifactor SEM was defined such that latent factors for each brain network were specified (two in the 2x2 analyses and four in the 4x5 analyses), but only one "general task" latent variable for all task states was specified (one latent task variable for both the 2x2 and 4x5 sets). The partial brain model tests the hypothesis that particular brain networks capture meaningful individual differences, but task-specific dimensions do not. If this were supported, it would imply that the only meaningful individual difference dimension is the brain network, and that t-fMRI does not add anything that is uniquely due to the particular task context. Such a pattern would be somewhat akin to suggesting that t-fMRI does not meaningfully capture between-subject variance beyond that which can be ascertained from resting-state fMRI. Given that t-fMRI explicitly tries to change neural activation patterns based on behavioral manipulations, we expect that it is highly unlikely that this is the best fitting model (see panel B of Figures 3 and 4).

*"Partial Task Model"*: This bifactor approach is the complement of the partial brain model such that one "general brain" latent factor was defined, but latent factors were specified for each of the task states (two in the 2x2 analyses and five in the 4x5 analyses). This bifactor model proposes that task contexts capture meaningful individual differences, but particular brain networks do not. If deemed the best fitting model, it would suggest that the task state is more impactful than functionally-defined brain networks, and that perhaps a majority of relevant information could be obtained via global whole-brain measures. Yet, there is precedent for observing region and network-specific brain-behavior correlations (Braver et al., 2010; Yarkoni & Braver, 2010), making this possibility less likely. For a schematic, see panel C of Figures 3 and 4.

**Figure 3. 2x2 Measurement Model Schematic**

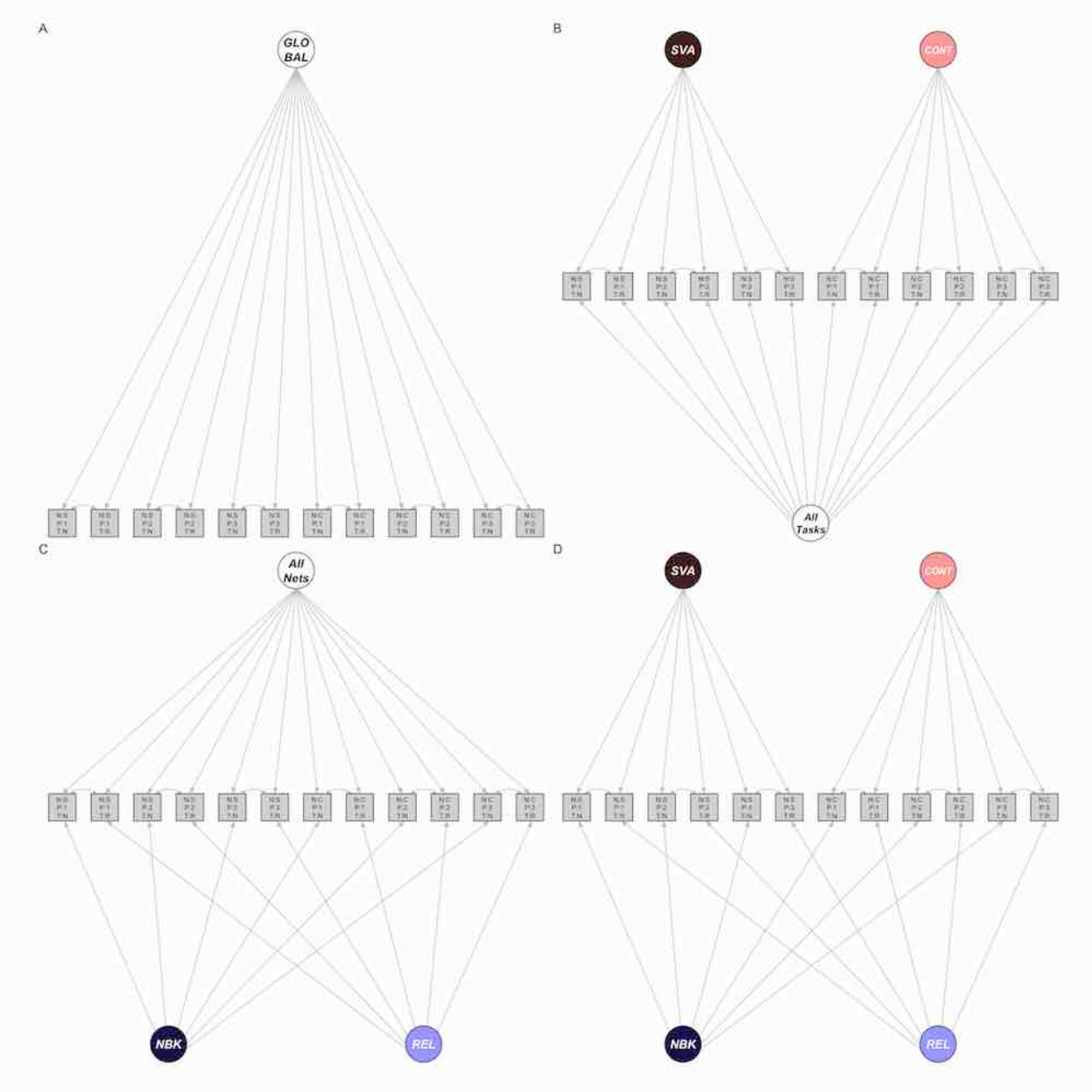
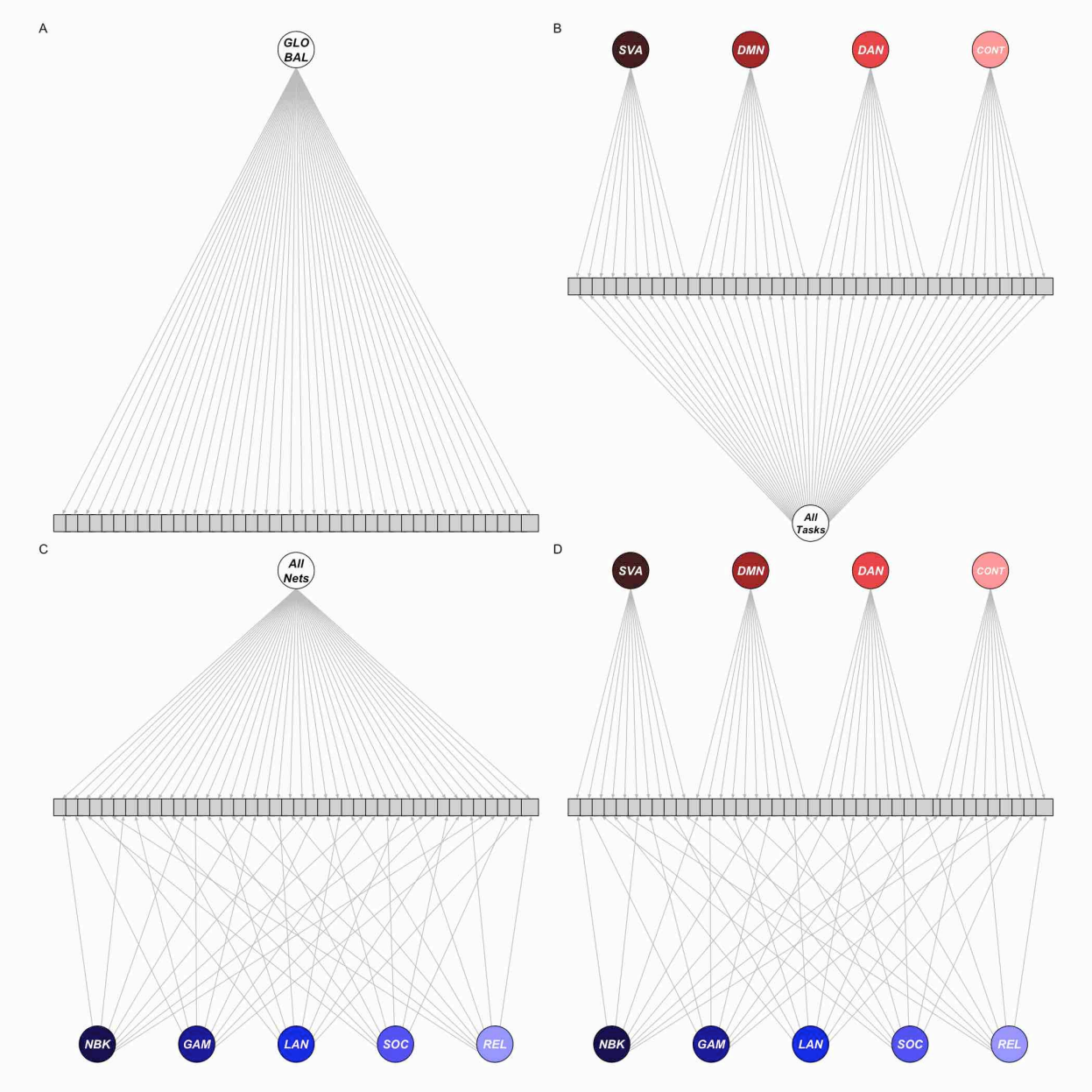


**Figure 4. 4x5 Measurement Model Schematic**




The focus of the first aim was to evaluate the four competing measurement models described above. For the second aim, the same procedures are repeated with the only change being the inclusion of the three outcome variables: List Sorting, Openness, and Grip Strength.

Three independent regression equations were specified per model such that each outcome was predicted by the defined latent variables.

All analyses were conducted in R (R Core Team, 2019); SEM models were constructed with the lavaan package (Rosseel, 2012) and all path diagrams were created with the semPlot package (Epskamp, 2015). All models were estimated using maximum likelihood with robust Huber-White standard errors (also known as a "sandwich variance estimator") in order to help protect against violations of multivariate normality (Kauermann & Carroll, 2001). Importantly, all models allowed the residuals of parcels to correlate with their corresponding parcel. For example, consider hypothetical parcel "A". Throughout the models described above, while the variance in parcel "A" is partitioned into an appropriate latent variable, there will still be some left over variability that cannot be explained by any of the latent variables. This residual variance is unique to that parcel, and could reflect any number of things; for example, a parcel's residual variance may be indicative of respiration patterns in that location of the brain. Since the same individuals completed multiple task paradigms (i.e., parcel "A" was measured during the N-back task, during the Relational Processing task, and so on), we therefore allowed the residual variances of each unique parcel to correlate (residual variance of parcel "A" in the N-back correlates with residual variance of parcel "A" in the Relational Processing task).

To reiterate, the first aim focused on the fit indices of the measurement models such that all models were pit against each other in a model comparison framework. Multiple fit indices (e.g., Root Mean Square Error of Approximation [RMSEA], Standardized Root Mean Square Residual [SRMR], Akaike Information Criterion [AIC], and Bayesian Information Criterion [BIC]) were examined to see how well each model's hypothesized covariance structure conforms

to the observed covariance structure. In all four of these fit measures, a lower value is better. Below briefly describes each fit index.

The RMSEA is a measure of absolute fit that compares the closeness between a hypothesized model and an ideal model, however it does penalize model complexity (Steiger, 1990). Conventional cutoffs of RMSEA are as follows: <.05 indicates very good fit, .05 - .08 indicates reasonable fit, and >.10 indicates poor fit. Additionally, 90% confidence intervals around the RMSEA are reported, with conventional wisdom suggesting that the upper bound of the 90% confidence interval should not exceed .10. Robust RMSEA values (including confidence intervals) are reported here because the current study utilized maximum likelihood estimation with robust standard errors.

The SRMR is another absolute fit measure, yet it is one where model complexity is not penalized. It indexes the standardized difference between observed correlations and hypothesized correlations; an SRMR <.10 is considered acceptable, and a SRMR equal to zero would be indicative of perfect fit.

The AICs and BICs are comparative fit indices that are especially useful for model comparisons. The model with the lowest AIC/BIC is considered the best-fitting model. Both AIC and BIC penalize for model complexity, however the BIC penalty is more severe, especially as the sample size increases. On the whole, AICs and BICs reported here converge in the same manner, but in the one or two instances where they yield differing results, we prioritize the BIC because it is more conservative. Furthermore, AIC and BIC are mathematically equivalent to k-fold cross-validation (Stone, 1977) and leave-one-out cross-validation (Shao, 1997),

respectively. The current study considers all fit indices; however, emphasis is placed on models with lowest AIC and BIC values.

When possible, determination of best competing model was accomplished via scaled chi-square difference ($\Delta\chi^2$) tests for nested model comparisons (Satorra & Bentler, 2001). Yet for models that have the same degrees of freedom, such as the partial brain, partial task models, and full bifactor models when no outcome measures are included; and partial brain and partial task models even when outcome measures are included, $\Delta\chi^2$ tests are not appropriate because they are not truly nested models and the difference in the difference in degrees of freedom is zero. Moreover, $\Delta\chi^2$ tests are known to be problematic. Most concerning for the current study, $\Delta\chi^2$ tests are meaningfully influenced by large sample sizes such that minute differences may become significant (Schermelleh-Engel, Moosbrugger, & Müller, 2003). We therefore take a holistic approach by considering all descriptive fit indices mentioned above in conjunction with $\Delta\chi^2$ tests when appropriate. In cases where fit measures tell differing stories, AICs and BICs were emphasized over the rest.

The second aim also considered fit statistics, but in this aim the focus was on the regression coefficients of latent variables predicting the outcome variables, as well as the variance of the outcomes that can be explained by the latent predictor variables. Regression coefficients reported are a result of both the manifest variables and latent variables being standardized and are thus denoted as "$b^*$". While a simple heuristic is to think of these as standardized regression coefficients from a normal linear regression, there are instances where the larger $b^*$ value is not significant, but a smaller $b^*$ value is significant. In all cases reported here, this is due to very large standard errors around the non-standardized regression coefficients (standard errors can be found in the full parameter estimate output on OSF). Importantly, while

significance of regression coefficients is described, priority is placed on the direction of association and overall magnitude so as not to overstate findings. Given that there are a large number of manifest variables and for all bifactor models (e.g., all models excluding the null models) there are two factor loadings per variable, factor weights are not reported here but can be found in the full parameter estimate output on OSF.

**Table 3. Number of Parcels Per Model**

| Model | Networks Included | Tasks Included | Latent Variable Description |
|---|---|---|---|
| 2x2 Analysis | | | |
| Null Model | SVA<br>CONT | N-back<br>Relational | Global Factor (50 parcels) |
| Partial Brain Model | SVA<br>CONT | N-back<br>Relational | SVA (24 parcels)<br>CONT (26 parcels)<br>Global Task Factor (50 parcels) |
| Partial Task Model | SVA<br>CONT | N-back<br>Relational | N-back (25 parcels)<br>Relational (25 parcels)<br>Global Brain Factor (50 parcels) |
| Full Bifactor Model | SVA<br>CONT | N-back<br>Relational | SVA (24 parcels)<br>CONT (26 parcels)<br>N-back (25 parcels)<br>Relational (25 parcels) |
| 4x5 Analysis | | | |
| Null Model | SVA<br>CONT<br>DAN<br>DMN | N-back<br>Relational<br>Gambling<br>Language<br>Social Cognition | Global Factor (320 parcels) |
| Partial Brain Model | SVA<br>CONT<br>DAN<br>DMN | N-back<br>Relational<br>Gambling<br>Language<br>Social Cognition | SVA (60 parcels)<br>CONT (65 parcels)<br>DAN (75 parcels)<br>DMN (120 parcels)<br>Global Task Factor (320 parcels) |
| Partial Task Model | SVA<br>CONT<br>DAN<br>DMN | N-back<br>Relational<br>Gambling<br>Language<br>Social Cognition | N-back (64 parcels)<br>Relational (64 parcels)<br>Gambling (64 parcels)<br>Language (64 parcels)<br>Social Cognition (64 parcels)<br>Global Brain Factor (320 parcels) |
| Full Bifactor Model | SVA<br>CONT<br>DAN<br>DMN | N-back<br>Relational<br>Gambling<br>Language<br>Social Cognition | SVA (60 parcels)<br>CONT (65 parcels)<br>DAN (75 parcels)<br>DMN (120 parcels)<br>N-back (64 parcels)<br>Relational (64 parcels)<br>Gambling (64 parcels)<br>Language (64 parcels)<br>Social Cognition (64 parcels) |

# Chapter 3: Results

## 3.1 Addressing Confounds Related to Family Structure of HCP Dataset

Before reporting the main findings from the current study, we address an important concern that SEM-related statistical inferences might be strongly impacted by the nested family structure of the HCP dataset (Van Essen et al., 2013). Though it is possible to define multi-level SEMs, it is sometimes challenging to do so, especially for bifactor SEMs, simply because the models often have issues converging. To ensure that using data from the entire HCP sample in a non-nested manner is a statistically valid approach, we examined this issue in terms of measurement invariance estimation. HCP participants were randomly assigned into two groups of unrelated individuals ($n_{group1}$ = 389 and $n_{group2}$ = 386; in cases where a family contained more than one individual, two of the family members were randomly chosen and randomly assigned to either group 1 or group 2 and all remaining family members were excluded). We then ran a measurement invariance procedure on the full bifactor 2x2 model with Schaefer 100 parcels. We defined a configural model where all parameters were freely estimated, and then a restricted model where all factor loadings were fixed to equal across the two groups. The idea is that if the configural model is measurably better than the equal loading model, then the models are not inherently similar across groups and thus combining all participants into one large group may be problematic for the current study procedures. Fit measures of the configural and equal loading models were extremely close, with perhaps the equal loading model being slightly better than the configural model. The AIC value favored the configural model by a very small margin ($AIC_{Configural}$ = 311,735 $AIC_{EqualLoadings}$ = 311,777), whereas the BIC favored the equal loadings model ($BIC_{Configural}$ = 314,025, $BIC_{EqualLoading}$ = 313,619). Moreover, the other fit measures were extremely close ($RMSEA_{Configural}$ = .103, $SRMR_{Configural}$ = .082; $RMSEA_{EqualLoading}$ = .101,

$SRMR_{EqualLoading}$ = .088). Taken together, these findings indicate that there are not meaningful differences between the configural and equal loadings models, and that allowing the factor loading parameters to be freely estimated in the configural model did not provide measurable benefits (in fact, these findings trend in the opposite direction, albeit only slightly). Given how close these two models appeared, we felt confident moving forward with the previously described set of analyses. However, we encourage future studies to carefully consider the hierarchical nature of these datasets.

## 3.2 Aim 1: 2x2 Analyses

The primary objective of this set of analyses was to determine which of the four competing measurement models best reflects the observed structure of t-fMRI BOLD data that were acquired during cognitive control-related task (N-back and Relational) conditions and looking within cognitive control-related brain networks (SVA/CON and Cont/FPN). The key hypothesis was that the full bifactor model would show better overall fit indices compared to the remaining three (null model, partial brain model, and partial task model), indicating that the separate brain networks and separate tasks contexts were all critical dimensions of individual differences. This hypothesis was clearly supported, and shown to be robust and consistent across the three different parcellations: Schaefer 300, Schaefer 100, and Gordon atlases. Table 4 shows that the full bifactor model was indeed the best fitting model across all fit indices for each of the three parcellations, although it is interesting that the 300 atlas has better fit indices compared to the 100 atlas. In fact, all three parcellations showed the same trend of the null model being the worst, the partial brain model being second worst, the partial task model being second best, and the full model being the best. Seeing as the degrees of freedom are identical for partial brain

models, partial task models, and full bifactor models (within a given atlas), $\Delta\chi^2$ tests were not run here.

**Table 4. Fit Indices of Aim 1 – 2x2 Analysis**

| Model | df | N Para-meters | AIC | BIC | RMSEA (90% CI) | SRMR |
|---|---|---|---|---|---|---|
| Dataset: Schaefer 100 | | | | | | |
| Null Model | 1,150 | 125 | 398,649.8 | 399,263.9 | .163 (.161, .165) | .233 |
| Partial Brain Model | 1,100 | 175 | 384,344.3 | 385,204.0 | .122 (.120, .124) | .146 |
| Partial Task Model | 1,100 | 175 | 382,481.5 | 383,341.2 | .115 (.113, .117) | .077 |
| Full Bifactor Model | 1,100 | 175 | 379,965.9 | 380,825.6 | .105 (.103, .106) | .078 |
| Dataset: Schaefer 300 | | | | | | |
| Null Model | 10,656 | 370 | 1,226,694.4 | 1,228,512.1 | .095 (.095, .096) | .196 |
| Partial Brain Model | 10,508 | 518 | 1,191,012.9 | 1,193,557.7 | .076 (.076, .077) | .165 |
| Partial Task Model | 10,508 | 518 | 1,184,351.6 | 1,186,896.4 | .072 (.071, .072) | .084 |
| Full Bifactor Model | 10,508 | 518 | 1,178,529.4 | 1,181,074.2 | .068 (.067, .069) | .079 |
| Dataset: Gordon | | | | | | |
| Null Model | 7,936 | 320 | 1,087,533.6 | 1,089,105.6 | .091 (.091, .092) | .177 |
| Partial Brain Model | 7,808 | 448 | 1,060,531.9 | 1,062,732.8 | .071 (.070, .072) | .142 |
| Partial Task Model | 7,808 | 448 | 1,056,679.7 | 1,058,880.6 | .067 (.067, .068) | .084 |
| Full Bifactor Model | 7,808 | 448 | 1,050,118.4 | 1,052,319.3 | .061 (.060, .062) | .072 |

## 3.3 Aim 1: 4x5 Analyses

The same analyses described above were repeated after including two additional brain networks (DMN and DAN) and three additional task contexts (Social Cognition, Language, and Gambling). Because of the increased complexity of this measurement model, this analysis was conducted only with the Schaefer 100 parcellation. The overall fit indices of the 4x5 analysis (Table 5) converged with findings from the 2x2 analysis (Table 4). The full bifactor model had the best fit indices, with the exception of the SRMR in the full bifactor model being just slightly higher than the partial task model (ΔSRMR = .007). Yet both AICs and BICs were lowest for the full model, supporting the same overall conclusions. Taken together, these findings strongly support the key hypothesis of the study: that cognitive tasks and brain networks are both

independent sources of individual differences even when tasks and networks are not cognitive control-specific.

**Table 5. Fit Indices of Aim 1 – 4x5 Analysis**

| Model | df | N Para-meters | AIC | BIC | RMSEA (90% CI) | SRMR |
|---|---|---|---|---|---|---|
| Dataset: Schaefer 100 | | | | | | |
| Null | 50,080 | 1,280 | 2,640,957.7 | 2,647,246.0 | .082 (.082, .082) | .173 |
| Partial Brain | 49,760 | 1,600 | 2,599,514.0 | 2,607,374.4 | .077 (.077, .077) | .164 |
| Partial Task | 49,760 | 1,600 | 2,505,306.7 | 2,513,167.1 | .063 (.063, .064) | .060 |
| Full | 49,760 | 1,600 | 2,500,084.8 | 2,507,945.2 | .063 (.062, .063) | .067 |

## 3.4 Aim 2: 2x2 Analyses

All four competing models were re-run with the inclusion of the three outcome variables (List Sorting, Openness, and Grip Strength) and defined regressions (i.e., measurement model and structural model) for each of the three parcellations. Fit indices are far more strongly influenced by the measurement model rather than the structural model of an SEM, and so unsurprisingly, the same model comparison results of the full bifactor being the best still held when including the regressions (fit statistics can be found for these 2x2 and 4x5 analyses in Supplement Tables 1 and 2, respectively). However, differing degrees allowed for $\Delta\chi^2$ tests to be run. Of note, there were cases where the $\Delta\chi^2$ statistic was negative. It is not appropriate to interpret findings in these cases, and therefore we default to relying on AIC and BIC values (note that it is possible to run a variant of a $\Delta\chi^2$ test that corrects for negative scaling factors [Satorra & Bentler, 2010], however implementation of this in the context of bifactor SEMs was quite challenging and not a fruitful endeavor). In all model comparisons that did not result in a negative scaling factor, results favored the more complex model (Tables 6 and 7). For the three

comparisons with a negative scaling factor, all fit indices support that the full bifactor model had the best fit (Table 4).

**Table 6. Chi-Squared Difference Tests – 2x2**

| Model 0 | Model 1 | $\Delta\chi^2$ | Δdf | *p*-value | Model To Keep? |
|---|---|---|---|---|---|
| Dataset: Schaefer 100 | | | | | |
| Null | Partial Brain | 4806.26 | 56 | < .001 | Model 1 |
| Null | Partial Task | 5417.91 | 56 | < .001 | Model 1 |
| Null | Full Bifactor | 7084.25 | 59 | < .001 | Model 1 |
| Partial Brain | Full Bifactor | *-1157.71* | 3 | *NA* | *NA* |
| Partial Task | Full Bifactor | *-644.57* | 3 | *NA* | *NA* |
| Dataset: Schaefer 300 | | | | | |
| Null | Partial Brain | 18673.7 | 154 | < .001 | Model 1 |
| Null | Partial Task | 18478.56 | 154 | < .001 | Model 1 |
| Null | Full Bifactor | 21752.46 | 157 | < .001 | Model 1 |
| Partial Brain | Full Bifactor | 705.29 | 3 | < .001 | Model 1 |
| Partial Task | Full Bifactor | *-3202.23* | 3 | *NA* | *NA* |
| Dataset: Gordon | | | | | |
| Null | Partial Brain | 14407.94 | 134 | < .001 | Model 1 |
| Null | Partial Task | 12415.58 | 134 | < .001 | Model 1 |
| Null | Full Bifactor | 15108.88 | 137 | < .001 | Model 1 |
| Partial Brain | Full Bifactor | 354.67 | 3 | < .001 | Model 1 |
| Partial Task | Full Bifactor | 3328.04 | 3 | < .001 | Model 1 |

*Note: When $\Delta\chi^2$ value is negative, it is inappropriate to interpret significance. Thus, models with negative $\Delta\chi^2$ values show NA for the last two columns. In these cases, we encourage readers to instead focus on AIC and BIC values for model comparisons.*

**Table 7. Chi-Squared Difference Tests – 4x5**

| Model 0 | Model 1 | $\Delta\chi^2$ | Δdf | *p*-value | Model To Keep? |
|---|---|---|---|---|---|
| Null | Partial Brain | 11059.26 | 332 | <.001 | Model 1 |
| Null | Partial Task | 38765.41 | 335 | <.001 | Model 1 |
| Null | Full | 42520.47 | 344 | <.001 | Model 1 |
| Partial Brain | Full | -10045.78 | 12 | *NA* | *NA* |
| Partial Task | Full | -1450.49 | 9 | *NA* | *NA* |

*Note: When $\Delta\chi^2$ value is negative, it is inappropriate to interpret significance. Thus, models with negative $\Delta\chi^2$ values show NA for the last two columns. In these cases, we encourage readers to instead focus on AIC and BIC values for model comparisons.*

Findings from Schaefer 100 and Schaefer 300 were markedly similar. See Figures 5 and 6, respectively, especially panel D. The N-back latent variable significantly positively predicted the List Sorting, after controlling for the three remaining variables ($b^*_{Schaefer\ 100} = .187$, $p < .001$; $b^*_{Schaefer\ 300} = .192$, $p < .001$). The Relational Processing latent task factor also positively predicted the List Sorting, but with a smaller effect size ($b^*_{Schaefer\ 100} = .075$, $p = .025$; $b^*_{Schaefer\ 300} = .062$, $p = .057$). While the relationships with the task factors aligned with our hypotheses, the brain network associations demonstrated some interesting trends. Interestingly, the Cont network (akin to the FPN in the Gordon parcels) did not significantly predict the List Sorting ($b^*_{Schaefer\ 100} = .052$, $p = .598$; $b^*_{Schaefer\ 300} = .094$, $p = .129$). Moreover, the SVA network (akin to CON in the Gordon parcels) did significantly predict List Sorting, however the relationship was negative ($b^*_{Schaefer\ 100} = -.094$, $p = .007$; $b^*_{Schaefer\ 300} = -.118$, $p < .001$). None of the latent variables from either Schaefer 100 or Schaefer 300 in the full bifactor model significantly predicted Openness or Grip Strength (see Supplemental Figures 1-4). When this process was repeated using the Gordon atlas (Figure 7), all coefficients were in the same direction as both Schaefer analyses, however the *p*-values were altered such that the N-back no longer reached significance ($b^*_{Gordon} = .200$, $p = .113$), while the FPN did ($b^*_{Gordon} = .088$, $p = .037$). The Relational Processing factor was still positively associated ($b^*_{Gordon} = .068$, $p = .042$) and the CON was showed the same significant negative association ($b^*_{Gordon} = -.085$, $p = .012$).

**Figure 5. 2x2 Structural Models to List Sorting with Schaefer 100**

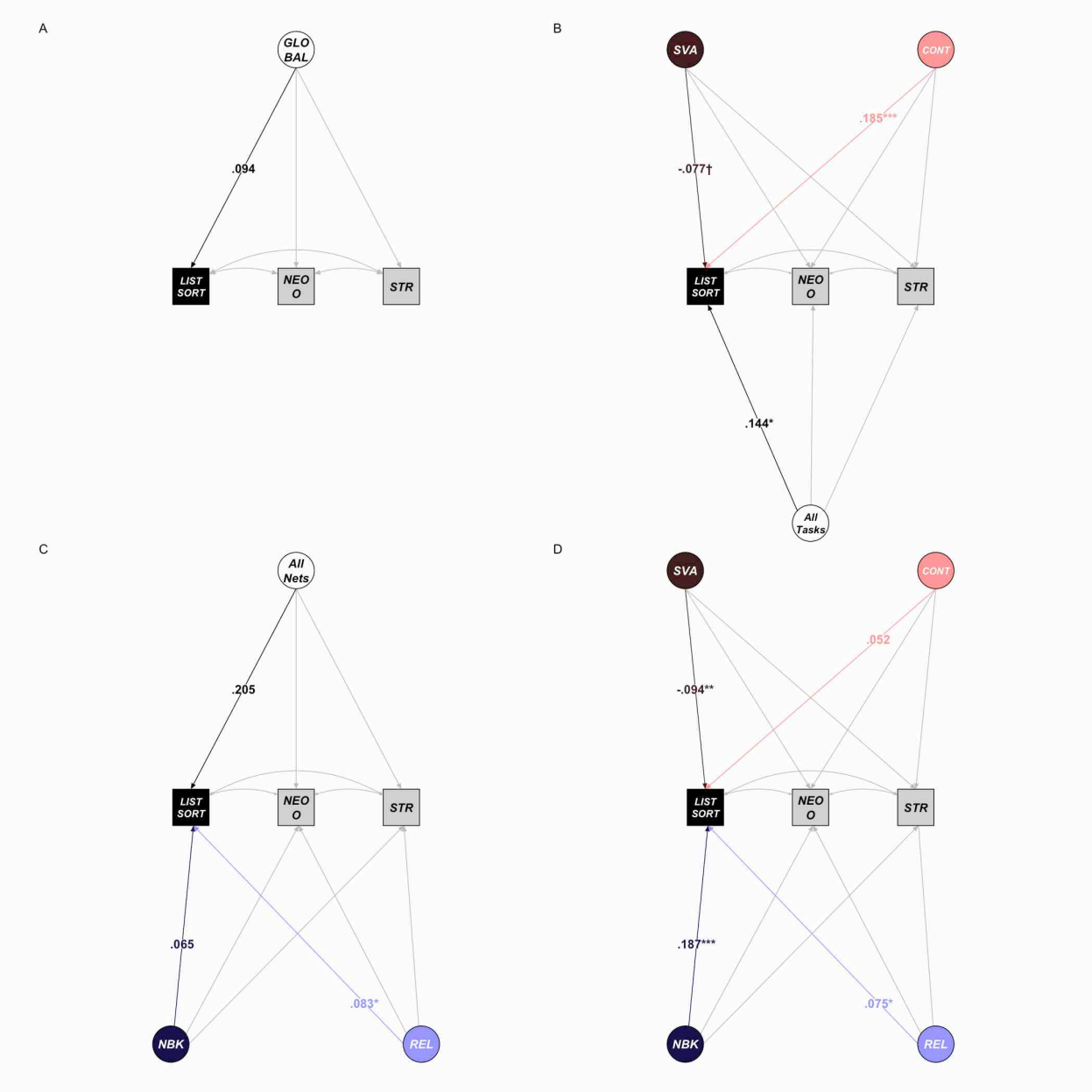


**Figure 6. 2x2 Structural Models to List Sorting with Schaefer 300**

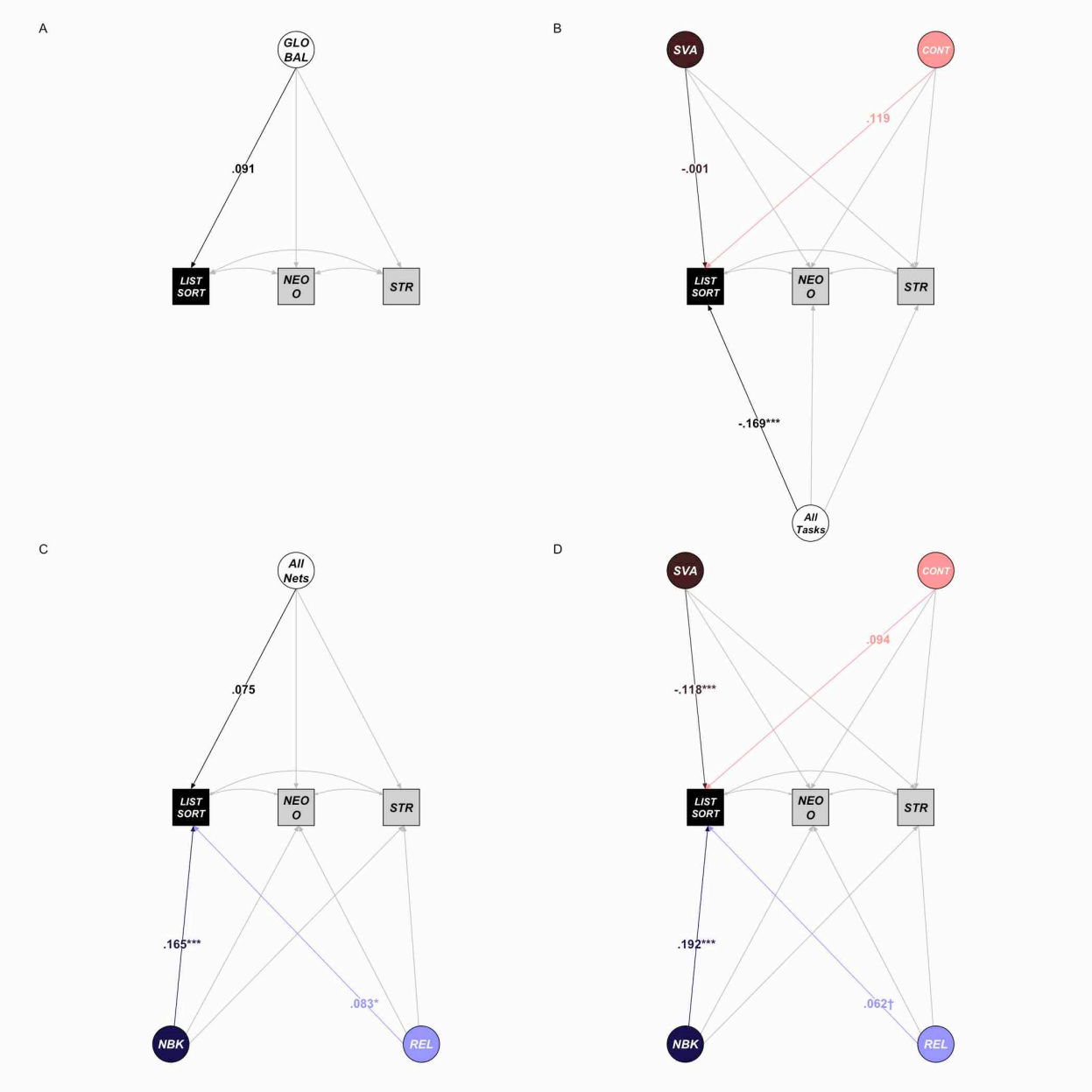
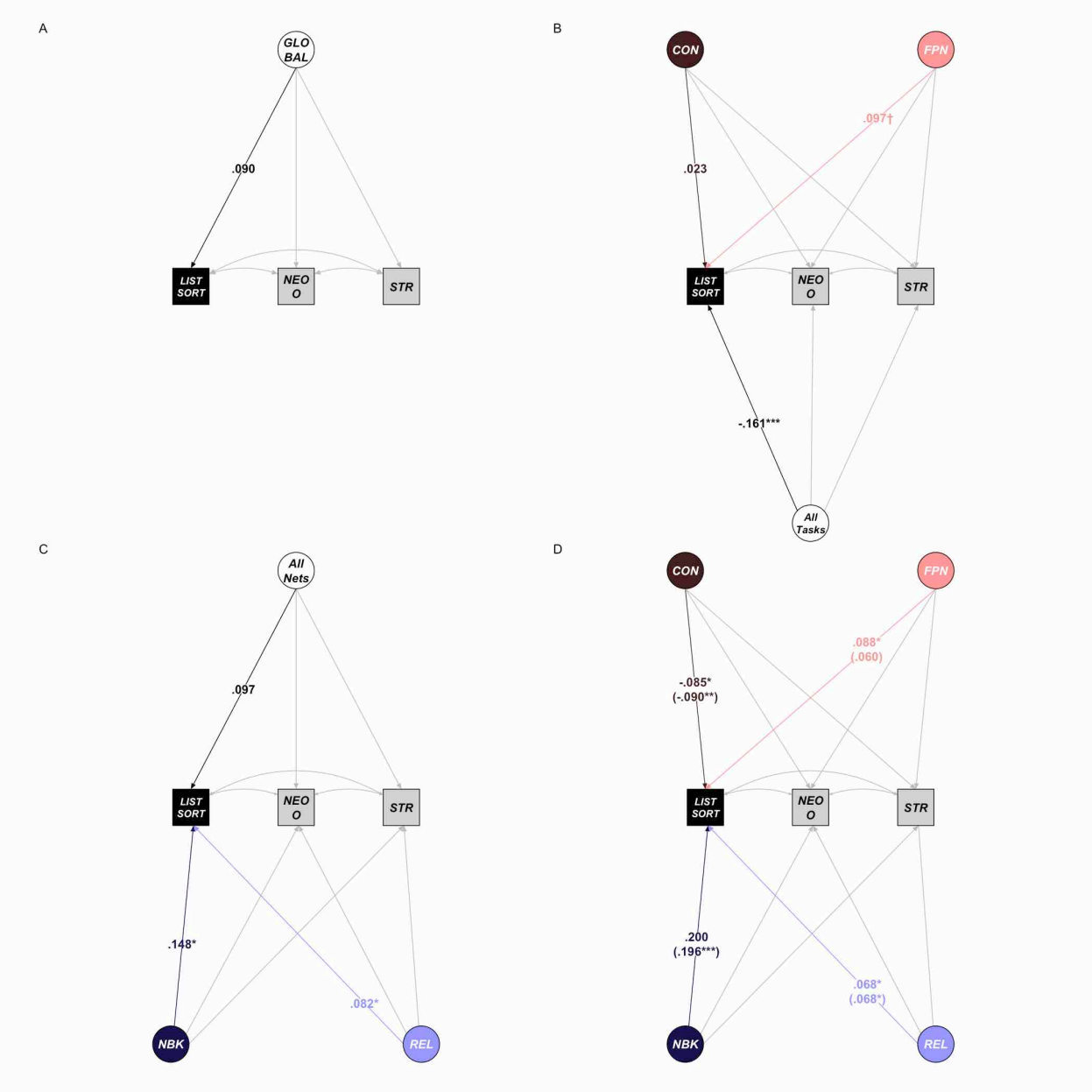


**Figure 7. 2x2 Structural Models to List Sorting with Gordon Atlas**




As just described, analysis of the Gordon parcellation yielded coefficients that were in the same direction as the coefficients in the Schaefer models, yet the statistical significance levels deviated for the N-back and the FPN/Cont (Schaefer: N-back significant, FPN/Cont not; Gordon: FPN/Cont significant, N-back not). Consequently, we explored whether allowing the residual

parcel variances to correlate impacted the regression findings (in Gordon models only). When the correlation of residual variances was not defined in the model, the associations to the List Sorting look more similar to the Schaefer parcels such that the N-back latent factor trends towards predicting the List Sorting ($b^*_{Gordon}$ = .196, $p$ = .085), while the FPN became no longer significant ($b^*_{Gordon}$ = .060, $p$ = .159), and the CON and Relational Processing coefficients showed the same associations as with the original Gordon atlas analysis and Schaefer analyses (CON: $b^*_{Gordon}$ = -.090, $p$ = .009, Relational Processing: $b^*_{Gordon}$ = .068, $p$ = .042). Lastly, to see if the trending of the N-back was impacted by the robust standard error procedure, we re-ran these analyses not including the residual variance correlations and without using the robust standard errors (standard maximum likelihood estimation). All regression coefficients remained the same, however without the robust standard errors, the N-back did reach significance ($b^*_{Gordon}$ = .196, $p$ < .001).

One explanation for the null results regarding Openness and Grip Strength may be that there is simply less variation in these measures than the List Sorting. To examine this descriptively, the coefficient of variation (standard deviation / mean, expressed as a percentage; CV%) was calculated for each of the three outcome measures. Surprisingly, the List Sorting had the lowest CV% whereas Openness had the highest: $CV\%_{ListSorting}$ = 12.89%, $CV\%_{Openness}$ = 21.81%, and $CV\%_{GripStrength}$ = 19.37%. Thus, the null Openness and Grip Strength findings cannot be attributed to there being less dispersion in these particular variables.

Although a lengthy and computationally-intensive process, the bifactor SEM approach used in these analyses did generally yield improved explanatory power of the List Sorting. Table 8 shows total explained variance in the outcome (e.g., List Sorting, Openness, Grip Strength) by all of latent variables defined in the model ($R^2$). In both the Schaefer 300 and Gordon atlases, the

$R^2$ values of the List Sorting in the full bifactor model were larger than the other three competing models (both of which have $R^2_{\text{List Sorting}} = .06$; Table 8). However, this pattern was not fully consistent in the Schaefer 100 atlas, at least descriptively, as here the highest List Sorting $R^2$ was noted in the partial brain model ($R^2 = .06$) whereas $R^2 = .05$ for both the partial task model and full bifactor model (see Table 8).

**Table 8. Variances and Variances Explained – 2x2**

| Model | Variable | Variance | SE | List Sorting % $R^2$ | Openness % $R^2$ | Strength % $R^2$ |
|---|---|---|---|---|---|---|
| Dataset: Schaefer 100 | | | | | | |
| Null Model | Global | .93 | 1.65 | .88% | .04% | .01% |
| Partial Brain Model | Global | 21.74 | 26.18 | | | |
| Partial Brain Model | CONT | 46.64** | 16.82 | 6.09% | .18% | .56% |
| Partial Brain Model | SVA | 57.38† | 30.15 | | | |
| Partial Task Model | Global | .11 | 2.03 | | | |
| Partial Task Model | NBK | 120.79*** | 16.98 | 5.31% | .33% | .21% |
| Partial Task Model | REL | 168.75*** | 19.76 | | | |
| Full Model | CONT | .06 | .20 | | | |
| Full Model | SVA | 74.61*** | 16.78 | 5.23% | .94% | .23% |
| Full Model | NBK | 63.94*** | 10.44 | | | |
| Full Model | REL | 166.33*** | 20.25 | | | |
| Dataset: Schaefer 300 | | | | | | |
| Null Model | Global | .41 | 1.03 | .82% | .13% | .01% |
| Partial Brain Model | Global | .04*** | .00 | | | |
| Partial Brain Model | CONT | .04 | .05 | 4.26% | 3.65% | .04% |
| Partial Brain Model | SVA | .04 | .07 | | | |
| Partial Task Model | Global | .02 | .06 | | | |
| Partial Task Model | NBK | 115.21*** | 21.12 | 3.97% | 1.71% | .04% |
| Partial Task Model | REL | 183.95*** | 27.43 | | | |
| Full Model | CONT | 1.87 | 2.22 | | | |
| Full Model | SVA | 85.26*** | 15.51 | 6.33% | 1.55% | .07% |
| Full Model | NBK | 88.03*** | 16.09 | | | |
| Full Model | REL | 175.15*** | 22.77 | | | |
| Dataset: Gordon | | | | | | |
| Null Model | Global | .11 | 1.52 | .81% | .09% | .00% |
| Partial Brain Model | Global | .01*** | .00 | | | |
| Partial Brain Model | FPN | 4.30 | 5.70 | 3.58% | 1.99% | .01% |
| Partial Brain Model | CON | .04 | .06 | | | |
| Partial Task Model | Global | .01 | .02 | | | |
| Partial Task Model | NBK | 39.19 | 27.95 | 3.8% | 1.39% | .02% |
| Partial Task Model | REL | 68.46* | 26.63 | | | |
| Full Model | FPN | 3.00 | 1.89 | | | |
| Full Model | CON | 152.25* | 66.83 | 5.95% | 1.13% | .04% |
| Full Model | NBK | 2.60 | 3.10 | | | |
| Full Model | REL | 101.54*** | 12.81 | | | |

*Note: Significance symbols reflect whether variance is significantly different from zero, with † indicating trending toward significance, * p < .05, ** p < .01, and *** p < .001. Confidence intervals around variances can be found in full parameter estimate outputs on OSF. SVA – Salience Ventral Attention network; CONT – Control network; FPN – Fronto-Parietal Network; CON – Cingulo-Opercular Network; NBK – N-back task; and REL – Relational Processing task.*

Additionally, Table 8 shows the (unstandardized) variances of each latent variable ($s^2$), standard errors around the $s^2$ estimates, and significance tests that ask if these variances are different from zero. Though nice to know, it is perhaps more informative to examine how the $s^2$ estimates change with each competing model (note that doing so is facilitated by the fact that in this study every manifest variable has the same underlying units rather than, say, a latent variable comprised of two brain measures and two behavioral measures). Interpretation of the variances is made easier when one considers how the latent variables are defined (readers may find the measurement models shown in Figure 3 to be particularly helpful here). For instance, take the $s^2$ for the Cont latent factor from the 2x2 Analysis in the Schaefer 100 (Table 8). In the partial brain model, Cont $s^2_{\text{PartialBrain}} = 46.64$, $se = 16.82$. In this model, Cont is interpreted as the variance shared across all parcels that have the "Control network" assignment, after controlling for any variance shared across all parcels (the Global latent factor). We can (descriptively) compare this variance to that in the full bifactor model; $s^2_{\text{FullBifactor}} = .06$, $se = .20$. Although this is a large decrease, the Cont latent factor in the full bifactor model reflects between-subject variability for all parcels with the "Control network" assignment, after controlling for any variance shared across parcels measured during the N-back task and any variance shared across parcels measured during the Relational task (N-back and Relational latent factors, respectively). This example highlights that when that when the observed variables are differentially organized into latent variables (e.g., latent variables defined different in each competing model), the variance captured by a given latent variable can markedly change.

In the Schaefer 100 dataset, the N-back and Relational variances both also decreased from the partial task model to the full bifactor model (some of their respective variances were then pulled into the SVA latent factor), although both still retained much more variability compared to the Cont (partial task model – $s^2_{\text{N-back}} = 120.79$, $s^2_{\text{Relatioal}} = 168.75$; full model – $s^2_{\text{N-back}} = 63.94$, $s^2_{\text{Relational}} = 166.33$). The SVA $s^2$ increased between the partial brain model and full bifactor model indicating that more SVA-unique variability was able to be pulled out when latent variables for tasks were defined (partial task model – $s^2_{\text{SVA}} = 57.38$; full model – $s^2_{\text{SVA}} = 74.61$). These same patterns mostly hold for the Schaefer 300 set, with the exception that the Cont did not demonstrate much variability in any of the four competing models. These patterns were largely seen in the Gordon dataset as well, with the notable exception that the N-back did not exhibit variances significantly different from zero in either the partial task or full bifactor models. Despite the fact that it is somewhat expected, generally, that $s^2$ estimates might decrease with the increasing numbers of latent variables (e.g., smaller variances in full bifactor model) because some of the variance will be partitioned into the newly added latent factor, these finding suggest that the full bifactor model still yields latent factors that capture between-subject variability (for full listing of all variances, please see Table 8).

As briefly mentioned earlier, the Openness factor was primarily chosen because it comes from the personality domain, rather than cognitive ability, and is known to correlate moderately with gF. However, one might expect different results if Openness was instead replaced with a cognitive ability measure of gF, like the PMAT. As such, the 2x2 analysis was re-run (on Schaefer 100 only) with the PMAT included instead of Openness to explore how findings may change. The overall fit indices still favored the full bifactor model over all others with the full bifactor model having the lowest AIC, BIC, and RMSEA (the partial task model had a slightly

lower SRMR, but the difference between the partial task model and full bifactor model was only .01; see Supplement Table 3 for full list of fit measures). Supplement Table 4 shows all regression results of latent variables predicting outcomes. The overall pattern of directionality of latent variables predicting List Sorting remained the same such that the N-back and Relational task latent factors significantly positively predicted List Sorting ($b^*_{N\text{-}back} = .187$, $p < .001$; $b^*_{Relational} = .073$, $p = .029$), the Cont showed no association to List Sorting ($b^*_{Cont} = .050$, $p = .543$), and the SVA showed a significant negative relationship ($b^*_{SVA} = -.093$, $p = .008$). Interestingly, relationships of these latent variables to the PMAT were nearly identical to those from of the latent variables to the List Sorting in terms of significance and direction of association ($b^*_{N\text{-}back} = .242$, $p < .001$; $b^*_{Relational} = .103$, $p = .001$; $b^*_{Cont} = .091$, $p = .520$; $b^*_{SVA} = -.131$, $p < .001$). Furthermore, the correlation between the PMAT and List Sorting in this analysis was .27 (compared to $r = .09$ for List Sorting and Openness). Given the strong association of the PMAT to List Sorting, it is not surprising that the significance and directionality of associations with the latent variables are mirrored.

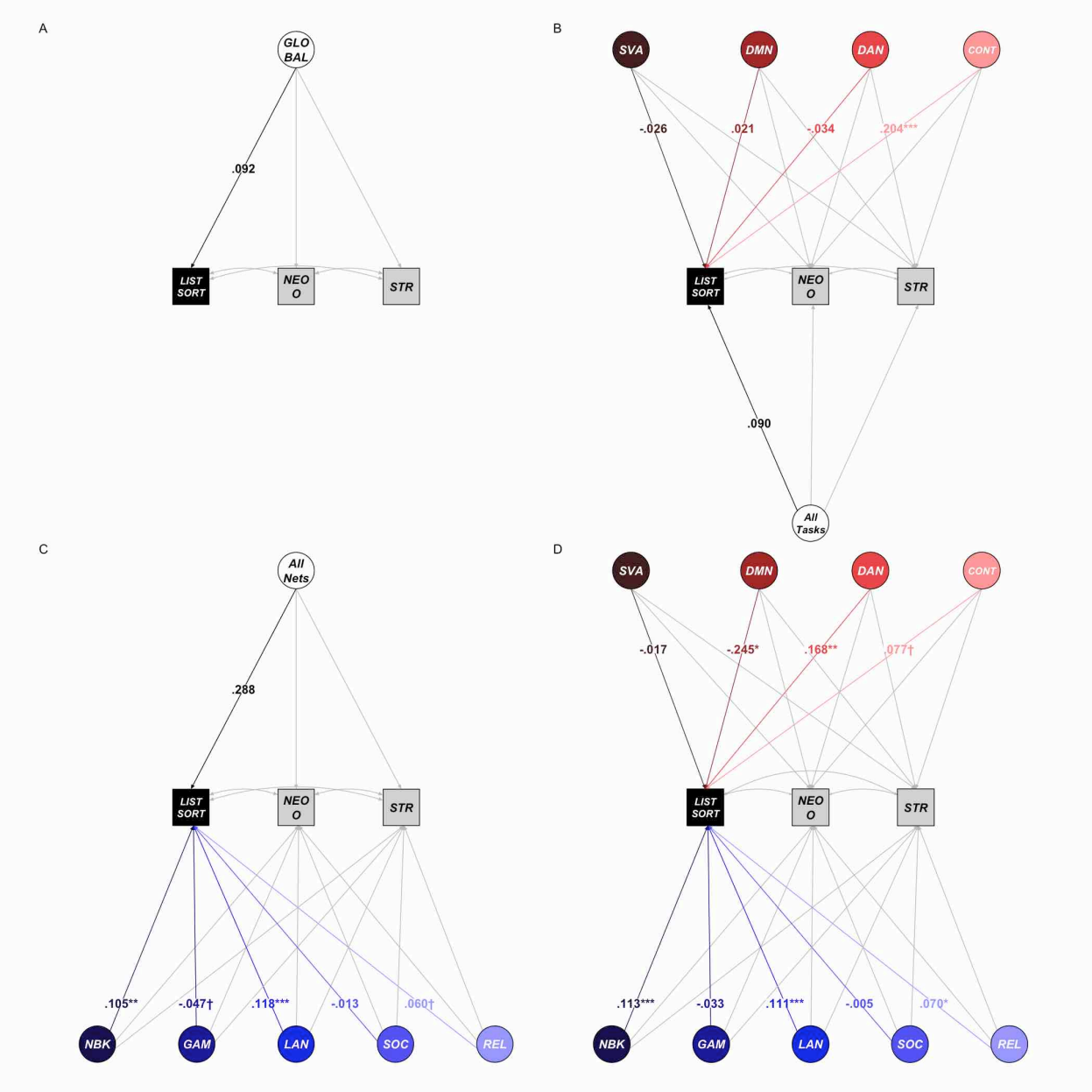
Interestingly, more variance was explained across the board in the PMAT than the List Sorting. For example, in the full bifactor model, 5.15% of the variance was explained in the List Sorting compared to 9.45% in the PMAT (see Supplement Table 5 for full list of $R^2$ values). Unlike previous analyses, however, this might be due to their being more variation in the PMAT compared to the List Sorting ($CV\%_{ListSorting} = 12.89\%$, $CV\%_{PMAT} = 27.60\%$). These findings suggest that perhaps the tasks and networks chosen for the 2x2 analyses are more closely related to a gF ability measure over a working memory measure, although please see the Discussion section for more on this particular topic.

Taken together, these results indicate that the overall associations across the three parcellation methods are indeed consistent, especially in regard to the directionality of all parameter estimates and model fits. The primary difference is that the strength of the associations between latent factors and outcomes in the Gordon atlas seem to be impacted by the inclusion of allowing residual variances to correlate. However, the direction of the associations, as well as the results of the model selection procedure (e.g., full bifactor model being the best model), align with the Schaefer models.

## 3.5 Aim 2: 4x5 Analyses

Overall fit indices of 4x5 SEM models that included outcomes supported the results of the measurement model (aim 1) outcomes. Given that the full model was the best fitting model, here we focus on the parameter estimates of relationships to the List Sorting from this model (Figure 8), however all coefficients for all four competing models can be found in Supplement Figures 5 and 6 for Openness and Grip Strength, respectively. See Table 7 for $\Delta\chi^2$ tests.

**Figure 8. 4x5 Structural Models to List Sorting with Schaefer 100**




In the 4x5 full bifactor model, the directions of associations between the original latent factors from the 2x2 (N-back, Relational Processing, SVA, and Cont) and the List Sorting persisted (see Figure 8, especially panel D). As before, the N-back and Relational Processing had positive significant associations with the N-back having a larger effect size than Relational

Processing ($b^* = .113$, $p = .001$ and $b^* = .070$, $p = .035$, respectively). For the brain networks, the direction of association remained however their significance levels changed. The positive relationship between the Cont network and List Sorting did not quite reach significance ($b^* = .077$, $p = .056$) and the SVA was negatively, although not significantly, related to List Sorting ($b^* = -.017$, $p = .693$). Similarly, the Gambling and Social Cognition tasks were negative but not significant predictors of the List Sorting task ($b^* = -.033$, $p = .234$ and $b^* = -.005$, $p = .873$, respectively). The remaining relationship between the Language task and List Sorting was significant and positive ($b^* = .111$, $p < .001$). The two additional brain networks were both significant predictors but in opposite directions such that the DMN had a negative association and the DAN had a positive association ($b^* = -.245$, $p = .011$ and $b^* = .168$, $p = .002$, respectively). These findings suggest, at least descriptively, that the same relationships from the 2x2 endure, even with the inclusion of higher-order but non-cognitive control-related brain networks and task contexts (note that one might want to formally test this via re-running the 4x5 model but fixing the regression parameters to the same coefficients from the 2x2 model, and comparing this fixed estimates model versus the free estimates model).

Note that the above is an example of the magnitude of $b^*$ not yielding the "most significant" coefficient (where $b^* = .111$, $p < .001$, but $b^* = -.245$, $p = .011$, for example). This is due to the larger standard errors around the regression coefficients. For the Language to List Sorting relationship, the *unstandardized* coefficient is b = .102, the standard error = .029, and thus z = 3.502. Yet for the DMN relationship to List Sorting, the unstandardized coefficient is b = -20.497, standard error = 8.040, and thus z = -2.549 (again, all unstandardized). Standardizing in a SEM framework is more nuanced than simply z-scoring because the standardization process includes the manifest variables and the latent variables. Although ideally it would be great to

obtain a $R^2$ value of each latent variable, this is not possible because the underlying assumption is that of local independence. This principle states that the latent variable explains why the manifest variables are related to one another – that is, the latent variable is a common predictor of each manifest variable (one might note that in path diagrams, the arrows point from the latent variable to the observed variables). One could get a $R^2$ for how much variance in each manifest variable is explained by a latent variable, although this is simply a conversion of each factor weight. We suggest readers use the standardized coefficients as a measure of relative magnitude and to place less emphasis on statistical significance, however noting that if a very large coefficient does not reach significance, that it is indicative of excessive error around that parameter. All unstandardized (and standardized) coefficients can be found in OSF parameter estimates outputs.

In the 2x2 analyses, none of the four latent variables predicted either Openness or Grip Strength (in either the original 2x2 analyses or in the 4x5 analyses). Yet interestingly, in the expanded 4x5 analysis, some of the added latent variables did predict these outcome variables. The relationship between both DMN and DAN to Openness mirrored their relationship to List Sorting such that they were in opposite directions, with the former trending toward significance ($b^* = -.154$, $p = .060$) and the latter reaching significance ($b^* = .208$, $p < .001$; see Supplement Figure 5). Finally, two of the nine latent variables significantly predicted the Grip Strength. The DMN showed a negative association ($b^* = -.095$, $p = .016$) as did the Social Cognition task ($b^* = -.093$, $p = .010$; see Supplement Figure 6).

Finally, the predictive utility was illustrated again in the 4x5 analyses such that the variance explained in the List Sorting was highest for the full bifactor model, although the partial task model was only slightly lower (see Table 9 for unstandardized variance estimates along with

$R^2$ values). The same trend was observed for variance explained in Openness, although all values were lower than those of the List Sorting. This trend was not observed for Grip Strength, however very little variance was explained by any model. Of note, more variance overall was explained in this set of 4x5 analyses compared to the 2x2 (e.g., for List Sorting in the Schaefer 100 atlas, $R^2_{2x2}$ = 5.23% and $R^2_{4x5}$ = 12.59%; see Tables 8 and 9). Furthermore, the variance estimates reported in Table 9 suggest that the task latent factors captured a lot of variability whereas the brain networks did not. As described in the Discussion section below, lower variances for the brain networks are not inherently problematic.

**Table 9. Variances and Variances Explained – 4x5**

| Model | Variable | Variance | SE | List Sorting *% $R^2$* | Openness *% $R^2$* | Strength *% $R^2$* |
|---|---|---|---|---|---|---|
| Dataset: Schaef100 | | | | | | |
| Null Model | Global | .07 | .19 | 0.86% | 0.08% | 0.01% |
| Partial Brain Model | Global | .04 | .15 | | | |
| Partial Brain Model | CONT | 51.34*** | 12.09 | | | |
| Partial Brain Model | SVA | 2.61 | 4.29 | 5.18% | 0.63% | 4.26% |
| Partial Brain Model | DMN | 2.06 | 7.96 | | | |
| Partial Brain Model | DAN | 2.17 | 4.94 | | | |
| Partial Task Model | Global | .02 | .04 | | | |
| Partial Task Model | NBK | 106.09*** | 16.70 | | | |
| Partial Task Model | REL | 171.54*** | 19.76 | 11.4% | 3.9% | 2.37% |
| Partial Task Model | GAM | 133.48*** | 20.74 | | | |
| Partial Task Model | SOC | 166.07*** | 25.62 | | | |
| Partial Task Model | LAN | 188.98*** | 23.48 | | | |
| Full Model | NBK | 111.18*** | 17.35 | | | |
| Full Model | REL | 162.00*** | 19.99 | | | |
| Full Model | GAM | 136.02*** | 21.51 | | | |
| Full Model | SOC | 160.19*** | 31.23 | | | |
| Full Model | LAN | 211.66*** | 23.58 | 12.59% | 8.28% | 3.08% |
| Full Model | CONT | .01** | .00 | | | |
| Full Model | SVA | .01 | .03 | | | |
| Full Model | DMN | .03* | .01 | | | |
| Full Model | DAN | .01* | .01 | | | |

*Note: Significance symbols reflect whether variance is significantly different from zero, with † indicating trending toward significance, * p < .05, ** p < .01, and *** p < .001. Confidence intervals around variances can be found in full parameter estimate outputs on OSF. SVA – Salience Ventral Attention network; CONT – Control network; DMN – Default Mode Network; DAN – Dorsal Attention Network; NBK – N-back task; REL – Relational Processing task; GAM – Gambling task; SOC – Social Cognition task; and LAN – Language task.*

# Chapter 4: Discussion

Findings from the current study support the notion that t-fMRI BOLD data contain separable sources of individual differences, and that isolating these sources through SEM approaches can be advantageous for enhancing explanatory power in brain-behavior relationships. This inference that the brain networks and task contexts are independent sources of individual differences is reasonably robust to parcellation algorithm. Furthermore, while significance levels did vary across parcellation method in the 2x2 procedures (although the Gordon models did ultimately match the Schaefer 100 and 300 models after some parameter tuning), and slightly between the 2x2 and 4x5 sets of analyses, the directionality of associations remained constant. Yet effect size ought to be prioritized over significance, especially when models are as incredibly large as the ones presented here. Thus, the fact that the direction of associations stayed the same and with reasonably similar effect sizes lends credibility to these findings.

We hypothesized that cognitive control is a domain in which both the brain networks and task contexts should be particularly vital dimensions of individual differences, with associated brain networks and task contexts showing meaningful relationships to an out-of-scanner measure of WM. While the first part of the hypothesis was supported, relationships between latent variables to WM were mixed. That the cognitive control-related tasks were positively linked to WM was not surprising, however the weaker/less reliable associations with the Cont/FPN, and the negative association of the SVA/CON were both unexpected. These oddities are discussed in further detail below (see *The Quiet FPN and Negative CON* subsection of this Discussion).

The expansion to the broader 4x5 analysis provided a stronger, but less hypothesis-driven extension of the study. Although the same full bifactor model was hypothesized to still be the

best fitting model in this expanded analysis, it was harder to make predictions regarding how inclusion of the additional networks and tasks would impact the findings. For instance, the three extra tasks (Social Cognition, Gambling, and Language tasks) all tap into higher-order cognitive processes, just like the N-back and Relational Processing tasks. If the underlying constructs were markedly similar across tasks, then the best fitting model of the 4x5 procedure may have been the partial brain model (independent brain networks, but one global task factor), and might have suggested that perhaps all 5 tasks were merely tapping an over-arching attentional state rather than narrower constructs. Likewise, evidence for a more general "task positive network" (Fox et al., 2005) would have manifested as similarities in between-subject variability across the four brain networks and thus the partial task model (independent task latent factors, but a global brain factor) might have had the best overall fit. And though the inclusion of the DMN may have complicated this slightly, one would have expected the DMN to show factor weights onto a general factor that were strong but negative. Though the partial task model was similar to the full model in terms of fit indices, the full model was selected because a) on the whole, fit measures were slightly better for the full model, especially AICs and BICs, and b) the full model better aligns with the larger literature that these brain networks are functionally distinct from one another.

In addition to the overall fits, interrogation of specific regressions in the 4x5 were also somewhat exploratory in that focus of the current study was to strategically examine if same relationships between cognitive control networks/tasks and WM found in the 2x2 model still remained present when expanding to the larger model, rather than articulating clear hypotheses for the remaining networks/tasks and their relationships to WM. However, interesting information was gleaned from these findings. First, the DMN was negatively associated with

nearly everything, which was reassuring and adds an element of construct validity to the results (Figure 8). Specifically, the DMN is known to deactivate when engaged in external cognitive tasks, which would suggest that the individual differences should scale with the strength of DMN deactivation, as was observed (Esposito et al., 2006; Raichle et al., 2001). Moreover, the strong relationship of the DAN to WM (Figure 8) was also somewhat anticipated given that the DAN has been shown to be sensitive to WM load (Majerus, Peters, Bouffier, Cowan, & Phillips, 2018). The DAN was also found to be associated with Openness ($b^* = .208$, $p < .001$; Supplement Figure 5), which was perhaps a bit more unexpected. However, the DAN and Cont/FPN are anatomically nearby (Vincent, Kahn, Snyder, Raichle, & Buckner, 2008) and perhaps inconsistencies in the literature, including but not limited to various parcellations, regarding the labeling of these networks could help explain the DAN to Openness relationship. Interestingly, the Language task was also strongly related to WM (Figure 8). This was especially surprising since the presented here analyses utilized the Story > Math contrast in order to account for WM since both the story and math conditions have equitable WM demands (Binder et al., 2011). However, language production has been linked to verbal WM (Acheson, Hamidi, Binder, & Postle, 2011), and the List Sorting is a verbal WM task. It would be even more unexpected if future studies find this relationship holds when including a relevant non-verbal outcome measure (e.g., a visuospatial task like the PMAT). Perhaps the most surprising and unpredicted finding was the significant, negative relationship of Social Cognition to Grip Strength ($b^* = -.093$, $p = .010$; Supplement Figure 6). On one hand, this could be indicative of a real relationship that ought to be explored in future studies. Conversely, it may not be particularly reliable, especially given that only about 3% of the variance in Grip Strength was explained across all latent variables (Table 9). As a control analysis, we obtained each individual's average activation

across all parcels (irrespective of network) measured during the Social Cognition task, examined the zero-order correlation of these averages to Grip Strength, and found a correlation of -.07 ($p$ = .029). The SEM procedure therefore only yielded a marginally larger effect size, suggesting that future studies may want to interrogate this relationship further. Yet given the small effect size and that not much variance in Grip Strength was explained, we caution against over-emphasizing this particular finding.

The variance estimates of the latent variables were particularly interesting in the 4x5 analyses in that they were notably large in the task factors and rather small for the brain factors (Table 9). Ideally, each latent factor would contain a lot of variability. However, the smaller variances reported here in the brain network factors are not inherently problematic. Of utmost importance is that the full bifactor model was better than the alternative models (Tables 4 and 5), indicating that hypothesized covariance patterns defined in the full model best matched the covariance patterns in the observed dataset. It would *not* be recommended to prefer the partial brain model over the full bifactor model, even though variances for the brain network factors were slightly larger in the partial brain model, because doing so would ignore the latent structure of the data. Moreover, one of the benefits of latent variable models is that latent variables are considered "error-free" and "perfectly reliable". While the network latent factors might not contain much variability, the amount that is there is more reflective of true score variability. As such, they can still be useful for subsequent analyses defined in the structural model.

On the whole, the current study supports the overarching hypothesis that the t-fMRI BOLD data contain separable dimensions of cognitive individual difference that can be partitioned into brain network and task context factors. The remainder of this section elaborates

on the implications of these results, the more surprising aspects these findings, some of challenges and limitations of these analyses, and potential future directions.

## 4.1 The Importance of Tasks

Cognitive psychologists and cognitive neuroscientists develop new task paradigms in order to tap into underlying cognitive constructs, which can then be used in t-fMRI experiments to better understand the neural mechanisms underlying such constructs. A common analysis and interpretation technique is to take the t-fMRI BOLD activation that was measured under a particular task, and correlate this activity with some behavioral outcome (either directly related to the task itself or a different out-of-scanner measure). A significant correlation is then interpreted as "individual differences in this brain region significantly relate to that behavior". However, this interpretation is somewhat misleading because it is not just individual differences in the brain region; rather, it is individual differences in a brain region *under a particular task context* that are related to the behavior. Put differently, individual differences in the brain region are "contaminated" by individual differences in the task (or vice versa), and thus it is a brain-by-task interaction. Perhaps one of the most important implications of the current study is that this interaction can be disentangled, and the findings presented here ultimately show that doing so is more reflective of the organization of t-fMRI variability than any of the other alternative models, including the null model (akin to the first component of a principal components analysis) or the partial brain model where the individual brain networks are distinguished from a global task component. That is, the task components add a critical element of between subject variability that cannot be found *only* in the brain regions. If anything, the fact that every analysis described here resulted in the partial task model having better overall fit indices than the partial brain

model suggests that the tasks may be even more important than particular brain regions, although, here, delineating both task contexts and brain networks was always best.

Moreover, the Human Connectome Project explicitly chose tasks designed to cover a breadth of functionality that could map onto distinct brain networks (Barch et al., 2013). As such, it is not surprising that the HCP tasks utilized in the current study contribute meaningful variability. The results of a similar analysis might be quite different if all tasks tapped a particular domain. For example, the Dual Mechanisms of Cognitive Control is an ongoing project that is scanning participants under four task paradigms, all of which broadly fit into the cognitive control domain (https://pages.wustl.edu/dualmechanisms). Examining how the findings presented here may replicate if using tasks from a more targeted construct is an interesting future direction. This topic is expanded upon below in the *Implications for the study of Cognitive Control* section of this Discussion.

While the current study explicitly shows the influence of task contexts as they pertain to t-fMRI studies, these findings this may also have vital repercussions for connectivity analyses. Functional connectivity and task activation are both crucial elements of healthy brain functioning, and an understanding of how they are intertwined will be critical in advancing cognitive neuroscience (see Cole, Ito, Bassett, & Sultz, 2016 for an interesting take on how these might be mathematically related). Though not discussed much here thus far, the majority of connectivity studies measure the BOLD signal during periods of awake rest. Yet if connectivity and activity are indeed enmeshed, then it holds that if the task setting is a critical dimension of in task activation analyses, then it may also play a substantial part in understanding individual differences in connectivity. In support of this claim, Finn et al., (2017) describe how connectivity can differ based on task state at both the between- and within-subject levels. They also found that

the ability to identify individual subjects based on connectivity patterns was notably worse in the rest state condition compared to the task conditions (Finn et al., 2017). These results, coupled with results of the current study, imply that consideration of task states should be critically important when trying to characterize the dimensions on which individuals differ.

## 4.2 The Quiet FPN and Negative CON

One particularly interesting finding of this investigation was that the Cont/FPN had a much weaker association with List Sorting than expected, including not reaching significance in the 2x2 analyses and only trending toward significance in the 4x5 analyses. This null result was surprising given that the relationship between Cont/FPN and cognitive control behaviors is well-documented in the literature, especially in that it is considered a flexible hub of connectivity supporting a variety of higher order functions (Cole et al., 2013). However, these findings may not be as paradoxical as they might seem. If the FPN is indeed a flexible hub, one might expect that it behaves somewhat like a relay station. For example, consider two separate cognitive tasks or goals. For the first task, information may enter the FPN, and due to the particular goal, the FPN will relay this information to an appropriate brain region for further processing (say, dorsolateral prefrontal cortex; DLPFC). In the second task, information enters the FPN, but since this goal is different from the first, perhaps the FPN relays this information to a different brain region for further processing (for example, perhaps the ventromedial prefrontal cortex; VMPFC). In this scenario, the FPN does indeed act as a flexible hub and one would expect the FPN to show strong increased task-related activity for these particular tasks. That is, while there may be between-subject variability, one might also predict that a truly flexible hub like the FPN would show marked between-task variability, perhaps even more so than between-subject variability. The findings presented here support this notion in that the current study statistically removes the

influence of between-task variance on brain network latent variables, and therefore there was little between-subject variance left over to be captured by a Cont/FPN latent factor. In the 2x2 analyses, the variance captured by the Cont network was not significantly different from zero for both the Schaefer 100 and Schaefer 300 atlases (where $s^2$ is the unstandardized variance of the Cont latent variable from the full bifactor model; $s^2_{Schaefer100} = .06$, $p = .782$; $s^2_{Schaefer300} = 1.87$, $p = .399$; Table 8). In the 4x5 analyses, the variance of the Cont network was significantly different from zero but it was still very minimal ($s^2 = .01$, $p = .002$; Table 9). Thus, while it may seem as though the null Cont/FPN finding is contradictory to the extensive literature, in fact the findings here might actually be a piece of converging evidence in favor of the Cont/FPN as a flexible hub of higher-order processing.

A similarly surprising finding was that the association between the SVA/CON and List Sorting was consistently negative. One potential explanation could be that participants were primarily engaged in reactive control which consists of a more late-onset conflict detection and performance monitoring system, as compared to proactive control which is thought to be more preparatory in nature (Braver, 2012). If an individual does not actively maintain task goals in their WM (as one might when using proactive control), then they instead must rely on stimulus-triggered reminders of the task demands or reactive control. As such, one might suspect that increased utilization of reactive control could be associated with decreased WM function. This is somewhat reminiscent of the Processing Efficiency Theory (Eysenck & Calvo, 1992) which posits that individuals with anxiety do not have as much of their WM capacity available due to their worries, which in turn leads to worse performance on WM tasks. In support of this, the anterior cingulate cortex, which is considered part of the SVA/CON and is thought to be involved with conflict detection, has been negatively associated with poorer WM performance

(Bunge, Ochsner, Desmond, Glover, & Gabrieli, 2001). A recent meta-analysis of decision-making tasks reports that the CON shows greater activity in a late-onset performance monitoring manner (Gratton et al., 2016), and thus the areas that comprise the SVA/CON may serve as a neural signature of reactive control, which in turn may therefore explain the negative associations between the SVA/CON and List Sorting.

The fact that we found the SVA/CON results to be somewhat surprising may be, unintentionally, due to inconsistent naming conventions. Some refer to these regions as the Salience network because they have been shown to be important for coordinating processes like attention and memory for stimuli that are particularly relevant, or "salient" to the task at hand (Menon & Uddin, 2010). Yet others, like both Schaefer and Gordon parcellations, use anatomical distinctions to define this network such as the Salience Ventral Attention network (SalVenAttn/SVA) or the Cingulo-Opercular network (CON), respectively. Thus, while the current study exploits the network neuroscience approach for the dimension reduction benefit, there is an important caveat in that using these networks can make it harder to connect to previous literature for this same reason. T-fMRI has a history of mostly exploring smaller nodes (e.g., dorsal anterior cingulate cortex) rather than the larger networks. This can make it quite difficult for researchers looking to previous work for hypothesis generation or those hoping to gain better understanding of their own findings via examining if there is any precedent for their findings (such as here with the negative relationship between SVA/CON and WM). Similarly, researchers may wind up searching for different key terms (i.e., searching for CON rather than Salience networks etc.) and inadvertently miss articles relevant to their research questions. Future individual differences t-fMRI studies that explicitly target network-level effects could help harmonize the overall literature. More broadly, we encourage future studies looking at

smaller ROIs to additionally include any network assignment (if applicable) one might see in the literature, and further support good open science practices to make it easier for researchers to compare ROIs and networks across studies (e.g., is this dorsolateral prefrontal cortex ROI from my study included in the FPN network of a different study?).

## 4.3 Parcellation Methods and Levels of Analysis

One of the most reassuring aspects of the current study is the general concordance across parcellation methods, especially in regard to the directionality of associations. As discussed above, the Gordon parcellation was a bit different from the Schaefer 100 and 300 atlases, although despite these differences the same overall patterns emerged. One possibility for some of the discrepancy is that the individual parcels of the Gordon atlas are more heterogeneous in size than the Schaefer parcels (Schaefer et al., 2018). This could potentially influence findings such that results could have been more heavily influenced by larger parcels in the Gordon atlas whereas the more equal parcel size of the Schaefer atlases would minimize this concern.

Interestingly, the Schaefer method allows researchers to decide how many parcels they would like (Schaefer et al., 2018). While this added flexibility is advantageous in allowing for more nuanced hypotheses, it can be very difficult for researchers to choose the appropriate granularity or dimensionality of parcellation, since the relative tradeoffs associated with this choice are unclear. The current study chose the 100 and 300 levels, the former to aid in dimensionality reduction and the latter to be comparable to the Gordon parcels ($n_{GordonParcels}$ = 333). We were careful to avoid repeating all analyses with all levels of parcels in order to minimize the likelihood of multiple comparisons concerns or, worse, falling prey to *p*-hacking. Yet there are certainly pros and cons to each level. From a classical test theory perspective, the benefit of obtaining more measures from an individual is that one is able to more accurately and

precisely capture a person's "true score" variance. Here, this would imply that one should prefer the Schaefer 300 over the Schaefer 100. Indeed, the precision of measurement is one potential explanation for why the fit indices, excluding AICs and BICs, of Schaefer 300 (and the Gordon parcels) were better than Schaefer 100 (Table 4).

The flip side of this network neuroscience advantage is the tradeoff of computational complexity. Indeed, researchers are constantly faced with computational complexity issues, and latent variable techniques like SEM are no exception. The more variables that are included in the SEM, the more difficult model estimation becomes. In fact, all of the Schaefer 300 models used up too much memory to be completed on a standard laptop and instead required using resources from the Washington University in St. Louis high-performance computing cluster, and the use of robust standard errors further increased the required computing resources. This reality led us to favor the Schaefer 100 when expanding to the 4x5 analyses (and again, all 4x5 analyses still required a computing cluster in order to run).

It is very possible that the results of the 2x2 or 4x5 analyses would be different if using a different level of parcellation, such as the 500 or 1000 atlases, although the degree to which they would differ is hard to characterize. Since the associations amongst latent factors and WM were consistent across the 100 and 300 atlases, we felt confident using the 100 atlas for the 4x5 analysis. If those relationships were not consistent, however, the interpretation of any of the analyses would have been far more cautious. Even still, far more emphasis was placed on the overall sign and effect size of the regression coefficient, rather than significance values, partially for this reason. The concordance across Schaefer 100 and Schaefer 300 seen here is also supported by Bolt, Nomi, Bainter, Cole, & Uddin (2019) who found that until one reaches the voxel level of analysis, the Schaefer parcels roughly yield the same inferences.

Thus far, the impact of parcellation methods have been one of the prominent methodological concerns of the current study. However, there is a different, often less explored level which is that of activation contrasts. It is very common in t-fMRI experiments to utilize some form of contrast comparison to tap an underlying behavior. For example, the N-back task used here had a 0-back task block (i.e., subjects should press a particular button when they see the target stimulus) and a 2-back task block (i.e., subjects should press a particular button when the stimulus shown on the screen is the same as the stimulus shown two trials before). Typically, researchers use these blocks to their advantage by creating contrasts such as the 2-back – 0-back blocks. In this example, the 0-back is not a particularly demanding WM task, whereas the 2-back has a much higher WM load, and so subtracting the 0-back activation from the 2-back activation allows researchers to target only activity that is exclusive to the increased WM load. This type of "narrow" contrast was used for each in-scanner task paradigm presented here. As such, it is possible that perhaps the individual differences captured by the tasks may have been weakened (e.g., the between-subject variance across tasks would have been more similar to one another) if a different, more liberal ("broad") activation contrast was used. For example, perhaps our findings would have differed if we had used a 2-back – average activation contrast or simply explored activation levels in the 2-back or 0-back blocks relative to just a common resting fixation (which is present in all tasks). However, these may be equally, if not more, problematic than the extreme contrast. Use of just a 2-back or just a 0-back condition, or an analogous block in a different task, often leads researchers with findings that are too coarse. Is the observed activity due to WM load or is it due to the participant engaging in any type of cognitively demanding task? The whole point of using contrasts is to enhance the signal relative to noise. Relatedly, while there are certainly times in which specific hypotheses about how an extreme

block would compare to average activity (2-back – avg) are theoretically justified, often these types of contrasts can be difficult to interpret. However, future studies may want to investigate the degree to which task-related individual differences vary based on activation contrasts.

## 4.4 Latent Variable Models and Neuroimaging – The Good, The Bad, and The Future

In earlier periods, cognitive psychology, and by extension cognitive neuroscience research, was mostly carried out independently of sub-fields focused on the study of individual differences (e.g., personality, intelligence etc.; Cronbach, 1957). However, in the last decade, questions related to individual differences have become more tightly integrated within the cognitive sciences. One of the unique aspects of this project is that the latent variable framework used here is optimized for the study of individual differences and is quite frequently used in domains where individual differences are at the forefront, yet it is still infrequently employed within neuroimaging research. The next section outlines some of the challenges and limitations faced by the application of latent variable techniques (e.g., SEM) to neuroimaging data in this study, as well as how utilization of these frameworks may be key in opening doors to new research question (for a more systematic and in-depth review of this topic, please see Cooper, Jackson, Barch, & Braver, 2019).

As described earlier, the SEM results provided some unexpected associations and potentially counter-intuitive results (see *The Quiet FPN and Negative CON* section of this Discussion), which is arguably the most exciting aspect of using this methodology. Yet one glaring limitation of the current study is that the fit indices, especially RMSEA and SRMR, are not as low as one might like. RMSEA values should ideally be less than .05, although .05-.08 are considered acceptable. RMSEA values of the best-fitting full bifactor models range from .061-

.105 (Tables 4 and 5), with no models scraping below the .05 ideal fit cutoff, including the lower 90% confidence intervals around the RMSEAs. Similarly, although SRMR indices met an acceptable cutoff, they could still be considered indicators of a mediocre fit. One driving force behind these lackluster fits is the large number of parameters being estimated (Tables 4 and 5). The sheer number of input/manifest variables makes these models somewhat daunting. The current study uses parcels as a middle ground between ROI and voxelwise approaches such that networks can be chosen based on *a priori* hypotheses, but do not need whole brain coverage, thus eliminating many more potential input variables. Future studies adopting latent variable methods may have more targeted hypotheses that would reduce the number of inputs and therefore simplify the model. For example, Bolt et al. (2018) took an SEM approach where they chose particular ROIs from a larger network (e.g., a ROI in right dorsolateral prefrontal cortex which had a network assignment of the FPN), and additionally they performed separate SEMs for each task. Consequently, the number of input variables were substantially decreased, and they report fit indices that traditionally fall into acceptable – very good ranges, especially in regard to SRMR (Bolt et al., 2018). While we hold that researchers can still use SEM for very large, complex models, we suggest that expectations be somewhat tempered as the number of inputs expands.

When using SEM, it is critical remember is that it is an inherently *disconfirmatory* procedure such that even with excellent fit measures, one can never truly prove that a model is the correct model. Instead, one can only eliminate worse models. In this regard, though the current study finds the full bifactor model to be the best-fitting, it is perhaps more important that we can eliminate the other three as potential choices. And perhaps even more importantly, SEM as applied in the current study is explicitly being used to test a hypothesis (note that it is possible

to perform exploratory SEMs but that is outside the scope of the current project). Conventional wisdom suggests that when fits are exceptionally close in model comparisons, researchers should rely on theory to guide their decisions (Kline, 2015). Taken together, the mediocre fits reported here still allowed us to make important headway: a) we were able to strike a balance between brain coverage and model complexity that still yielded reasonable fits, and b) we were able to eliminate worse competing models via taking a holistic approach to fit indices and incorporating ideas from previous literature.

An additional methodological limitation of the current study, outside of the topic of fit measures, is that the HCP contains twin and sibship pairs, yet the analyses presented here do not account for this hierarchical family structure. Further, rather than limit the participants to only those that are unrelated (making the sample size roughly half of what we report here), we instead chose to use all participants for several reasons. Most notably, the large number of parameters being estimated in these SEM models requires exceedingly large sample sizes (Kline, 2015). Although we did perform a measurement invariance procedure to try to mitigate these concerns, we fully acknowledge that it would be much better to account for the nested structure of these data. Although it is possible to conduct hierarchical SEMs (in the lavaan R package, only 2 levels are allowed as of version 0.6-5; Rosseel, 2012), it can be quite difficult to overcome convergence issues. Furthermore, typically even more participants are required for hierarchical analyses than traditional analyses. Estimating power with SEM is not quite straightforward and requires fairly involved bootstrapping procedures, and though we do not report on power for the models presented here, we air on the side of assuming we are underpowered. Big data projects like the ongoing ABCD study will likely have enough power to detect if the findings presented here replicate once the family structure is properly accounted for.

A final statistical concern, which can be problematic across most areas of psychology and neuroscience, is the notion of overfitting. Though it has been argued that psychology ought to switch its focus more towards prediction, as opposed to explanation (Yarkoni & Westfall, 2017), the SEM approach applied here was not optimized for prediction. Rather, the objective of most SEMs is to get the most accurate parameter estimates. As an analogy, consider standardized tests like the Graduate Records Examination (GRE). Development of standardized testing has very strong roots in latent variable techniques like SEM, often using a related method known as item-response theory. Yet the goal of the GRE is to get the most precise measure of an individual's abilities – not to see if a matched participant would get a similar score.

Despite the goals of SEM not being particularly geared toward predictions, there are some tools one could use to feel more confident in how their SEM would hold in a prediction-based framework. One possibility is to use a cross-validation approach. One could split the dataset into a larger "training set", develop the model, and then define that same model with the same (fixed) parameter estimates but using the remaining "testing set" and examining the model fits. This can be done a number of times and ultimately results from the testing sets would be examined to see how well the original model held. If the various iterations of the testing set demonstrated good fits, one might infer that the SEM would hold out-of-sample. To the authors' knowledge, this procedure is not performed much in the SEM literature, if at all. One reason is, again, the need for large samples which is of course impeded if the original dataset is split into training and testing sets. If this wanted to be accomplished with the current set of models, the HCP is likely too small of a dataset (note that it might be possible where the data are split into an 80:20 training/testing sets, yet even this requires averaging across iterations or folds, which can introduce additional complexities). However, the ongoing ABCD study, which aims to include

~11,500 kids (Casey et al., 2018), might have large enough sample sizes where this procedure could be accomplished. Although, it is worth noting that AIC and BIC values mathematically converge with cross-validation studies such that a model with the lowest AIC and BIC would perform best in k-fold cross-validation (Stone, 1977) and leave-one-out cross-validation (Shao, 1997) processes, respectively. As such, while future studies might want to go through this process more thoroughly, it may not yield results that are any more informative than that which can already ascertained from the AIC and BIC values of traditional SEM model outputs.

Another possibility in addressing this overfitting concern may be to take a measurement invariance procedure, similar to the one described on the family structure here, but where each group comes from a different dataset. Of course, this would require very similar tasks, similar preprocessing, and may be quite challenging. However, many of the large-scale neuroimaging studies are following open science practices and encouraging of data sharing (Poldrack & Gorgolewski, 2014). Though it would take careful consideration, it may be a possibility for future studies.

Most immediately, however, future studies may want to address overfitting via employing regularization, which essentially penalizes models based on complexity. In the context of SEM, this can be applied in a frequentist or Bayesian manner (see Jacobucci & Grimm, 2018). In the frequentist form of regularization, the notion is that parameters with small estimates are essentially set to zero so as to minimize the contribution of parameters that may not be as critical. In contrast, regularization in Bayesian SEM essentially allows parameters with small estimates to still be estimated, but with very limited variability so as to a get a more accurate representation of the parameter (Jacobucci & Grimm, 2018). Either way, regularization

procedures could help balance the same over-fitting concerns without going through an iterative cross-validation process that compromises sample sizes.

## 4.5 Implications for the study of Cognitive Control

While the preceding section focused on limitations and future directions in regard to methodology, this last section addresses the limitations and implications the current findings have for improving our understanding cognitive control.

As briefly described above, one limitation of the current study is that the HCP tasks were chosen for breadth across domains, rather than depth of one or two domains. Researchers particularly interested in a given domain may instead opt to include multiple task paradigms of one construct or multiple paradigms of a small number of highly interconnected constructs. The currently on-going Dual Mechanisms of Cognitive Control study is doing just that. This project is scanning participants under four task paradigms that are all thought to be part of cognitive control, including the Stroop, AX-CPT, Cued Task Switching, and Sternberg tasks (https://pages.wustl.edu/dualmechanisms). If these four tasks were used in the current study, rather than the 5 HCP tasks, one might instead expect the partial brain model to be favored. Since the four tasks are tapping the same underlying construct, it may therefore be better to consider items from each task as multiple measurements of the same latent factor (i.e., a global task factor). This pattern of findings would be interpreted as indicating that the broader domain of cognitive control is an important source of between-subjects variability, but that each task in and of itself is not a meaningful dimension of cognitive individual difference. Yet a recent study tried to use structural equation modeling on multiple task paradigms to create an executive function latent variable but were unsuccessful in their endeavor (Rey-Mermet, Gade, Souza, Bastian, & Oberauer, 2019). Therefore, questions regarding the relative utility of studies

involving multiple tasks tapping a single psychological construct versus individual paradigms tapping multiple constructs remain an open area of research.

An additional limitation of the current study relates to the outcome dimensions. We chose three outcome measures that we hypothesized would have varied degrees of relatedness to cognitive control. Much the same way we do not use a single parcel as a predictor in the current study and instead define latent variables comprised of multiple indicator parcels, ideally outcome variables would be latent variables that consisted of at least three or more measures of that particular domain. Since the outcome would also be latent, it should only reflect "true score" variance in the construct and be free of random error. In turn, this should help strengthen any true brain-behavior relationships. For an example of this, please see Example 4 of Cooper, Jackson, Barch, & Braver (2019). As such, future studies ought to use outcome variables from which they can define a latent outcome construct. Yet we advise researchers to first examine the measurement model(s) of just the outcomes to ensure it is suitable to be absorbed into a larger model.

Cognitive control has been linked to a variety of clinical disorders, so much so that it is even a construct of interest in the NIMH RDoC Matrix. Yet the treatment of psychological conditions is notoriously difficult. There is a plethora of reasons for why this may be the case, one of which might be that previous t-fMRI research of the behaviors most impaired in patient populations do not delineate the influences of tasks versus the influences of brain networks. Consider what might happen if the analyses described in the current study were repeated on a cohort with clinical impairments known to be associated to cognitive control. Given the findings here that cognitive control-related tasks seem to be their own sources of individual differences, it follows that clinical dysfunctions linked to cognitive control might also exhibit task-related

individual differences. This might manifest similarly to what was described here where the full model is best, or perhaps the variability is best captured by specific task latent variables and a global brain latent variable (e.g., partial task model). In either scenario, researchers may want to consider closer interrogation of those task dimensions and their relationships to therapeutics that help moderate a person's behavior within an environment. That is, the individual differences captured by the task state might be critical for predicting which individuals would benefit most from cognitive behavioral therapy versus mindfulness meditation versus exposure therapy etc. Further examinations of the influences of tasks and brain networks, especially in regard to cognitive control, are thus warranted and may be a key pathway toward precision medicine efforts.

Whereas the HCP cohort included healthy young adults, future studies may want to investigate if brain networks and task contexts are crucial dimensions at different stages of development and decline. Not only is SEM particularly well-suited for longitudinal data analyses, but the currently ongoing ABCD study might be the ideal dataset on which to examine hypotheses relating to how the sources of individual differences may change over the course of development. Since ABCD has a very large target sample size of ~11,500, SEM methods described here can be easily ported with the added benefit of increased statistical power. As of this study, wave 1 of ABCD (9-10 year old children) has been publicly released, however the entire project will follow these children for 10 years with imaging assessments roughly every 2 years (for more information, please see https://abcdstudy.org/). One possible hypothesis that could very feasibly be tested with these ABCD data is that independent brain networks dimensions might not capture individual differences as well as task contexts in very young children, as experiences play a large role in forming brain architecture. As they grow up,

however, the best fitting models might ultimately evolve to have both brain networks and task contexts be important individual difference dimensions.

## 4.6 Conclusions

There seems to be a growing appreciation that human behavior is rarely categorical in nature; there are a variety of dimensions on which people differ. The difficulty, however, is in picking and choosing these dimensions. For example, it is easy to make the assumption that people differ in their BOLD activation patterns, but it is much more difficult to determine if researchers ought to focus on continuums at the level of individual brain networks versus larger whole brain patterns or even smaller ROIs. The current study serves as a proof-of-concept, highlighting that applications of modern psychometric frameworks like latent variable modeling in conjunction with big data neuroimaging projects can feasibly help researchers in this endeavor. The analysis techniques described here can be modified to accommodate more targeted hypotheses and even different datasets. Adoption of these methods along with further psychometric considerations and refinements specific to neuroimaging could set the stage for exciting future research, especially in regard to disentangling brain and task related variability. Benefits include increasing explanatory power of brain-behavior relationships in a psychometrically sound way, as well as statistical opportunities to reveal new insights that might otherwise be overshadowed by the coupling of brain activation and task contexts (e.g., the CON's negative relationship with WM). We hope future studies targeting brain-behavior relationships, will continue to explore how these relationships may differ across task contexts both within and outside of the cognitive control domain.

# References


Acheson, D. J., Hamidi, M., Binder, J. R., & Postle, B. R. (2011). A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience*, *23*(6), 1358–1367. http://doi.org/10.1162/jocn.2010.21519

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. http://doi.org/10.1037/0033-2909.131.1.30

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., et al. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *Neuroimage*, *80*(C), 169–189. http://doi.org/10.1016/j.neuroimage.2013.05.033

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, *19*(3), 354–369. http://doi.org/10.1177/1073191112446655

Binder, J. R., Gross, W. L., Allendorfer, J. B., Bonilha, L., Chapin, J., Edwards, J. C., et al. (2011). Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *Neuroimage*, *54*(2), 1465–1475. http://doi.org/10.1016/j.neuroimage.2010.09.048

Bollen, K. A. (1989). Structural Equations with Latent Variables. Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/9781118619179

Bolt, T., Nomi, J. S., Bainter, S. A., Cole, M. W., & Uddin, L. Q. (2019). The situation or the person? Individual and task-evoked differences in BOLD activity. *Human Brain Mapping*, *40*(10), 2943–2954. http://doi.org/10.1002/hbm.24570

Bolt, T., Prince, E. B., Nomi, J. S., Messinger, D., Llabre, M. M., & Uddin, L. Q. (2018). Combining region- and network-level brain-behavior relationships in a structural equation model. *Neuroimage*, *165*, 158–169. http://doi.org/10.1016/j.neuroimage.2017.10.007

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in lisrel maximum likelihood estimation. *Psychometrika*, *50*(2), 229–242. http://doi.org/10.1007/BF02294248

Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn Affect Behav Neurosci*, *7*(4), 356–366. http://doi.org/10.3758/cabn.7.4.356

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. http://doi.org/10.1016/j.tics.2011.12.010

Braver, T. S., & Barch, D. M. (2006). Extracting core components of cognitive control. *Trends in Cognitive Sciences*, *10*(12), 529–532. http://doi.org/10.1016/j.tics.2006.10.006

Braver, T. S., Cole, M. W., & Yarkoni, T. (2010). Vive les differences! Individual variation in neural mechanisms of executive control. *Current Opinion in Neurobiology*, 1–9. http://doi.org/10.1016/j.conb.2010.03.002

Bunge, S. A., Ochsner, K. N., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (2001). Prefrontal regions involved in keeping information in and out of mind. *Brain : a Journal of Neurology*, *124*(Pt 10), 2074–2086. http://doi.org/10.1093/brain/124.10.2074

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., et al. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54. http://doi.org/10.1016/j.dcn.2018.03.001

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns. *Neuroimage*, *12*(3), 314–325. http://doi.org/10.1006/nimg.2000.0612

Chiew, K. S., & Braver, T. S. (2011). Positive affect versus reward: emotional and motivational influences on cognitive control. *Frontiers in Psychology*, *2*, 279. http://doi.org/10.3389/fpsyg.2011.00279

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *Neuroimage*, *37*(1), 343–360. http://doi.org/10.1016/j.neuroimage.2007.03.071

Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and Task-Evoked Network Architectures of the Human Brain. *Neuron*, *83*(1), 238–251. http://doi.org/10.1016/j.neuron.2014.05.014

Cole, M. W., Ito, T., Bassett, D. S., & Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, *19*(12), 1718–1726. http://doi.org/10.1038/nn.4406

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, *16*(9), 1348–1355. http://doi.org/10.1038/nn.3470

Cooper, S. R., Jackson, J. J., Barch, D. M., & Braver, T. S. (2019). Neuroimaging of individual differences_ A latent variable modeling perspective. *Neuroscience & Biobehavioral Reviews*, *98*, 29–46. http://doi.org/10.1016/j.neubiorev.2018.12.022

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain., *3*(3), 201–215. http://doi.org/10.1038/nrn755

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. http://doi.org/10.1037/h0043943

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the Hemodynamic Responses to Reward and Punishment in the Striatum. *Journal of Neurophysiology*, *84*(6), 3072–3077. http://doi.org/10.1152/jn.2000.84.6.3072

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, *104*(26), 11073–11078. http://doi.org/10.1073/pnas.0704320104

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., et al. (2006). A core system for the implementation of task sets. *Neuron*, *50*(5), 799–812. http://doi.org/10.1016/j.neuron.2006.04.031

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: the organization of goal-directed behavior. *Cognitive Psychology*, *30*(3), 257–303. http://doi.org/10.1006/cogp.1996.0008

Epskamp, S. (2015). semPlot: Unified Visualizations of Structural Equation Models. *Structural Equation Modeling: a Multidisciplinary Journal*, *22*(3), 474–483. http://doi.org/10.1080/10705511.2014.937847

Esposito, F., Bertolino, A., Scarabino, T., Latorre, V., Blasi, G., Popolizio, T., et al. (2006). Independent component model of the default-mode brain function: Assessing the impact of active thinking. *Brain Research Bulletin*, *70*(4-6), 263–269. http://doi.org/10.1016/j.brainresbull.2006.06.012

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and Performance: The Processing Efficiency Theory. *Cognition and Emotion*, *6*(6), 409–434. http://doi.org/10.1080/02699939208409696

Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *Neuroimage*, 1–12. http://doi.org/10.1016/j.neuroimage.2017.03.064

Firth, J., Stubbs, B., Vancampfort, D., Firth, J. A., Large, M., Rosenbaum, S., et al. (2018). Grip Strength Is Associated With Cognitive Performance in Schizophrenia and the General Population: A UK Biobank Study of 476559 Participants. *Schizophrenia Bulletin*, *44*(4), 728–736. http://doi.org/10.1093/schbul/sby034

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, *103*(26), 10046–10051. http://doi.org/10.1073/pnas.0604187103

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678. http://doi.org/10.1073/pnas.0504136102

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. http://doi.org/10.1038/nature18933

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*, *80*(C), 105–124. http://doi.org/10.1016/j.neuroimage.2013.04.127

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology*, *84*(4), 537–552. http://doi.org/10.1037/0022-0663.84.4.537

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2016). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex*, *26*(1), 288–303. http://doi.org/10.1093/cercor/bhu239

Gratton, C., Neta, M., Sun, H., Ploran, E. J., Schlaggar, B. L., Wheeler, M. E., et al. (2016). Distinct Stages of Moment-to-Moment Processing in the Cinguloopercular and Frontoparietal Networks. *Cerebral Cortex*, bhw092–15. http://doi.org/10.1093/cercor/bhw092

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–322. http://doi.org/10.1038/nn1014

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*(1), 253–258. http://doi.org/10.1073/pnas.0135058100

Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F., & Weinberger, D. R. (2002). The Amygdala Response to Emotional Stimuli: A Comparison of Faces and Scenes. *Neuroimage*, *17*(1), 317–323. http://doi.org/10.1006/nimg.2002.1179

Jacobucci, R., & Grimm, K. J. (2018). Comparison of Frequentist and Bayesian Regularization in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 639–649. http://doi.org/10.1080/10705511.2017.1410822

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671. http://doi.org/10.3758/BF03196323

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169–183. http://doi.org/10.1037//0096-3445.130.2.169

Kauermann, G., & Carroll, R. J. (2001). A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *Journal of the American Statistical Association*, *96*(456), 1387–1396.

Kline, R. B. (2015). Principles and Practice of Structural Equation Modeling (4 ed.). New York: Guilford Publications.

Lerman-Sinkoff, D. B., Sui, J., Rachakonda, S., Kandala, S., Calhoun, V. D., & Barch, D. M. (2017). Multimodal neural correlates of cognitive control in the Human Connectome Project. *Neuroimage*, *163*, 41–54. http://doi.org/10.1016/j.neuroimage.2017.08.081

Majerus, S., Peters, F., Bouffier, M., Cowan, N., & Phillips, C. (2018). The Dorsal Attention Network Reflects Both Encoding Load and Top-down Control during Working Memory. *Journal of Cognitive Neuroscience*, *30*(2), 144–159. http://doi.org/10.1162/jocn_a_01195

Marek, S., & Dosenbach, N. U. F. (2018). The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. *Dialogues in Clinical Neuroscience*, *20*(2), 133–140.

McCrae, R. R., & Costa, P. T., Jr. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, *36*(3), 587–596. http://doi.org/10.1016/S0191-8869(03)00118-1

Medaglia, J. D., Lynall, M.-E., & Bassett, D. S. (2015). Cognitive Network Neuroscience. *Journal of Cognitive Neuroscience*, *27*(8), 1471–1491. http://doi.org/10.1162/jocn_a_00810

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure & Function*, *214*(5-6), 655–667. http://doi.org/10.1007/s00429-010-0262-0

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. http://doi.org/10.1006/cogp.1999.0734

Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: data sharing in neuroimaging. *Nature Neuroscience*, *17*(11), 1510–1517. http://doi.org/10.1038/nn.3818

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional Network Organization of the Human Brain. *Neuron*, *72*(4), 665–678. http://doi.org/10.1016/j.neuron.2011.09.006

Power, J. D., Schlaggar, B. L., & Petersen, S. E. (2014). Studying Brain Organization via Spontaneous fMRI Signal. *Neuron*, *84*(4), 681–696. http://doi.org/10.1016/j.neuron.2014.09.007

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682. http://doi.org/10.1073/pnas.98.2.676

Redick, T. S. (2014). Cognitive control in context: Working memory capacity and proactive control. *Acta Psychologica*, *145*, 1–9. http://doi.org/10.1016/j.actpsy.2013.10.010

Reuben, D. B., Magasi, S., McCreath, H. E., Bohannon, R. W., Wang, Y. C., Bubela, D. J., et al. (2013). Motor assessment using the NIH Toolbox. *Neurology*, *80*(Issue 11, Supplement 3), S65–S75. http://doi.org/10.1212/WNL.0b013e3182872e01

Rey-Mermet, A., Gade, M., Souza, A. S., Bastian, von, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General*, *148*(8), 1335–1372. http://doi.org/10.1037/xge0000593

Richmond, L. L., Redick, T. S., & Braver, T. S. (2015). Remembering to prepare: The benefits (and costs) of high working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1764–1777. http://doi.org/10.1037/xlm0000122

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. http://doi.org/10.18637/jss.v048.i02

Sadaghiani, S., & D'Esposito, M. (2015). Functional Characterization of the Cingulo-Opercular Network in the Maintenance of Tonic Alertness. *Cerebral Cortex*, *25*(9), 2763–2773. http://doi.org/10.1093/cercor/bhu072

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. http://doi.org/10.1007/BF02296192

Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika*, *75*(2), 243–248. http://doi.org/10.1007/s11336-009-9135-y

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: evidence from pupillometry and event-related fMRI. *Neuroimage*, *37*(3), 1017–1031. http://doi.org/10.1016/j.neuroimage.2007.04.066

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., et al. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. http://doi.org/10.1093/cercor/bhx179

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research*, *8*(2), 23–74.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *Journal of Neuroscience*, *27*(9), 2349–2356. http://doi.org/10.1523/JNEUROSCI.5587-06.2007

Seitzman, B. A., Gratton, C., Marek, S., Raut, R. V., Dosenbach, N. U., Schlaggar, B. L., et al. (2018). A set of functionally-defined brain regions with improved representation of the subcortex and cerebellum. *bioRxiv*, 450452.

Shafto, M. A., Tyler, L. K., Dixon, M., Taylor, J. R., Rowe, J. B., Cusack, R., et al. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, *14*, 204. http://doi.org/10.1186/s12883-014-0204-1

Shao, J. (1997). AN ASYMPTOTIC THEORY FOR LINEAR MODEL SELECTION. *Statistica Sinica*, *7*(2), 221–242.

Sporns, O., & Bassett, D. S. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. http://doi.org/10.1038/nn.4502

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, *25*(2), 173–180. http://doi.org/10.1207/s15327906mbr2502_4

Stone, M. (1977). An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 44–47.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases

of Middle and Old Age. *PLoS Medicine*, *12*(3), e1001779. http://doi.org/10.1371/journal.pmed.1001779

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. http://doi.org/10.1152/jn.00338.2011

Tulsky, D. S., Carlozzi, N., Chiaravalloti, N. D., Beaumont, J. L., Kisala, P. A., Mungas, D., et al. (2014). NIH Toolbox Cognition Battery (NIHTB-CB): list sorting test to measure working memory. *Journal of the International Neuropsychological Society : JINS*, *20*(6), 599–610.

Ugurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A. T., Duarte-Carvajalino, J. M., et al. (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage*, *80*(C), 80–104. http://doi.org/10.1016/j.neuroimage.2013.05.012

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., & Consortium, F. T. W.-M. H. (2013). The WU-Minn Human Connectome Project: An overview. *Neuroimage*, *80*(C), 62–79. http://doi.org/10.1016/j.neuroimage.2013.05.041

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of Neurophysiology*, *100*(6), 3328–3342. http://doi.org/10.1152/jn.90355.2008

Viscogliosi, G., Di Bernardo, M. G., Ettorre, E., & Chiriac, I. M. (2017). Handgrip Strength Predicts Longitudinal Changes in Clock Drawing Test Performance. An Observational Study in a Sample of Older Non-Demented Adults. *The Journal of Nutrition, Health & Aging*, *21*(5), 593–596. http://doi.org/10.1007/s12603-016-0816-9

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., et al. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Developmental Cognitive Neuroscience*, *32*, 4–7. http://doi.org/10.1016/j.dcn.2017.10.002

Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding Animate Agents: Distinct Roles for the Social Network and Mirror System. *Psychological Science*, *18*(6), 469–474. http://doi.org/10.1111/j.1467-9280.2007.01923.x

Wig, G. S., Laumann, T. O., & Petersen, S. E. (2014). An approach for parcellating human cortical areas using resting-state correlations. *Neuroimage*, *93*(Part 2), 276–291. http://doi.org/10.1016/j.neuroimage.2013.07.035

Yarkoni, T., & Braver, T. S. (2010). Cognitive Neuroscience Approaches to Individual Differences in Working Memory and Executive Control: Conceptual and Methodological Issues. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of Individual Differences in Cognition* (pp. 87–107). New York, NY: Springer New York. http://doi.org/10.1007/978-1-4419-1210-7_6

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, *11*, 174569161769339–23. http://doi.org/10.1177/1745691617693393

# Appendix

**Supplement Table 1**

| Model | df | N Para-meters | AIC | BIC | RMSEA (90% CI) | SRMR |
|---|---|---|---|---|---|---|
| Dataset: Schaefer 100 | | | | | | |
| Null | 1,297 | 134 | 422,101.2 | 422,759.5 | .154 (.152, .156) | .221 |
| Partial Brain | 1,241 | 190 | 407,771.3 | 408,704.7 | .116 (.114, .117) | .138 |
| Partial Task | 1,241 | 190 | 405,899.2 | 406,832.7 | .109 (.107, .111) | .073 |
| Full | 1,238 | 193 | 403,382.3 | 404,330.4 | .099 (.098, .101) | .075 |
| Dataset: Schaefer 300 | | | | | | |
| Null | 11,097 | 379 | 1,250,145.5 | 1,252,007.5 | .093 (.093, .094) | .192 |
| Partial Brain | 10,943 | 533 | 1,214,431.2 | 1,217,049.7 | .075 (.074, .075) | .162 |
| Partial Task | 10,943 | 533 | 1,207,765.8 | 1,210,384.3 | .071 (.070, .071) | .083 |
| Full | 10,940 | 536 | 1,201,927.5 | 1,204,560.7 | .067 (.066, .067) | .078 |
| Dataset: Gordon | | | | | | |
| Null | 8,317 | 329 | 1,110,985.2 | 1,112,601.5 | .089 (.089, .090) | .173 |
| Partial Brain | 8,183 | 463 | 1,083,958.3 | 1,086,232.9 | .069 (.069, .070) | .139 |
| Partial Task | 8,183 | 463 | 1,080,101.9 | 1,082,376.5 | .066 (.065, .067) | .083 |
| Full | 8,180 | 466 | 1,073,526.0 | 1,075,815.4 | .060 (.059, .060) | .071 |

**Supplement Table 2**

| Model | df | N Para-meters | AIC | BIC | RMSEA (90% CI) | SRMR |
|---|---|---|---|---|---|---|
| Dataset: Schaefer 100 | | | | | | |
| Null | 51,037 | 1,289 | 2,664,408.8 | 2,670,741.4 | .081 (.081, .081) | .171 |
| Partial Brain | 50,705 | 1,621 | 2,622,922.2 | 2,630,885.7 | .076 (.076, .077) | .163 |
| Partial Task | 50,702 | 1,624 | 2,528,635.5 | 2,536,613.8 | .063 (.063, .063) | .060 |
| Full | 50,693 | 1,633 | 2,523,389.5 | 2,531,412.0 | .062 (.062, .062) | .067 |

**Supplement Table 3**

| Model | df | N Para-meters | AIC | BIC | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| Dataset: Schaefer 100 | | | | | | |
| Null | 1,150 | 125 | 398,649.8 | 399,263.9 | .163 (.161, .165) | .233 |
| Partial Brain | 1,100 | 175 | 384,344.3 | 385,204.0 | .122 (.120, .124) | .146 |
| Partial Task | 1,100 | 175 | 382,481.5 | 383,341.2 | .115 (.113, .117) | .077 |
| Full | 1,100 | 175 | 379,965.9 | 380,825.6 | .105 (.103, .106) | .078 |

**Supplement Table 4**

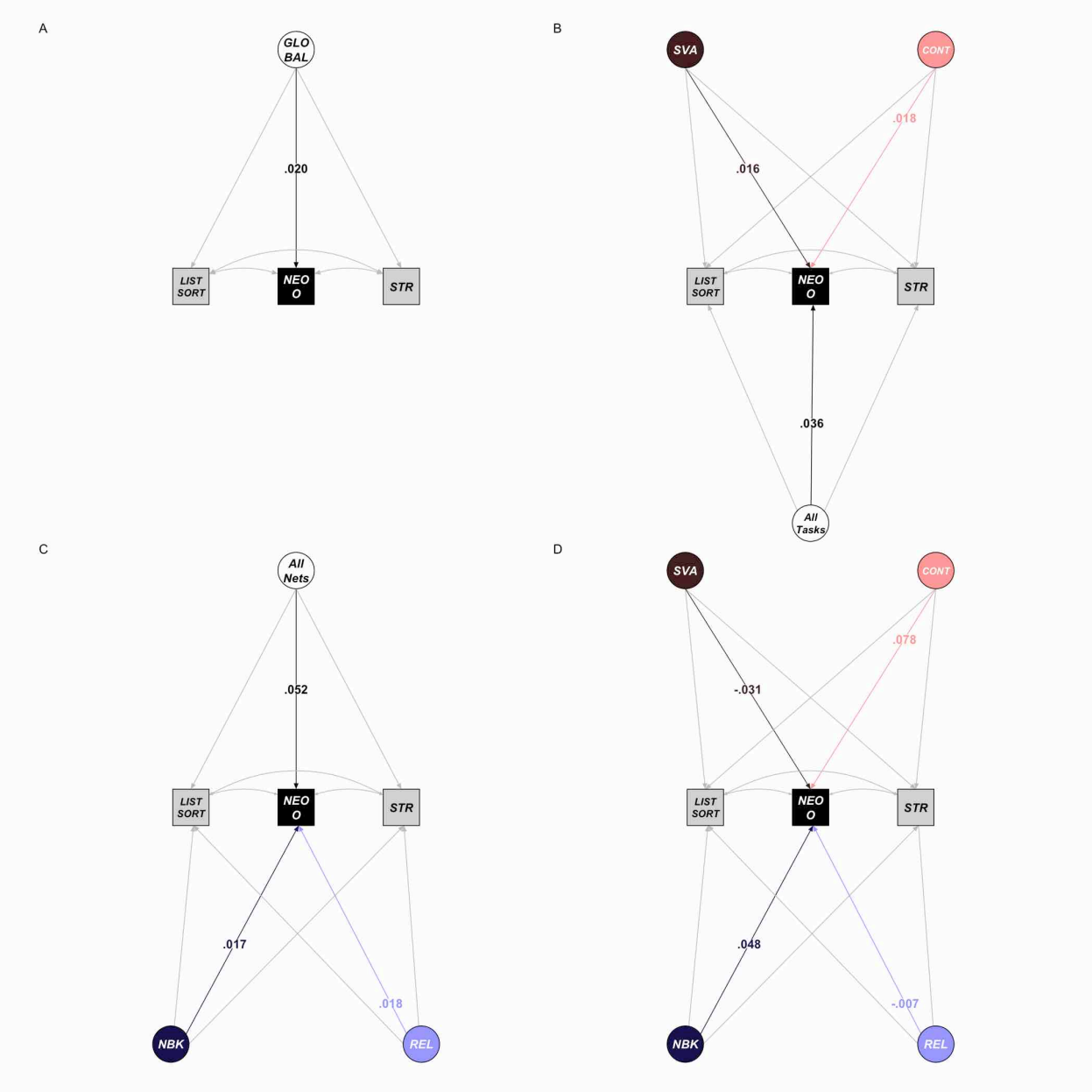
| Latent Variable | List Sorting $b^*$ | PMAT $b^*$ | Grip Strength $b^*$ |
|---|---|---|---|
| SVA | -.093** | -.131*** | -.027 |
| CONT | .050 | .091 | -.014 |
| NBK | .187*** | .242*** | .034 |
| REL | .073* | .103** | .016 |

*Note: Significance symbols reflect if regression coefficient is significantly different from zero: * p < .05, ** p < .01, and *** p < .001. SVA - Salience Ventral Attention network; CONT – Control network; NBK – N-back task; REL – Relational Processing task. **

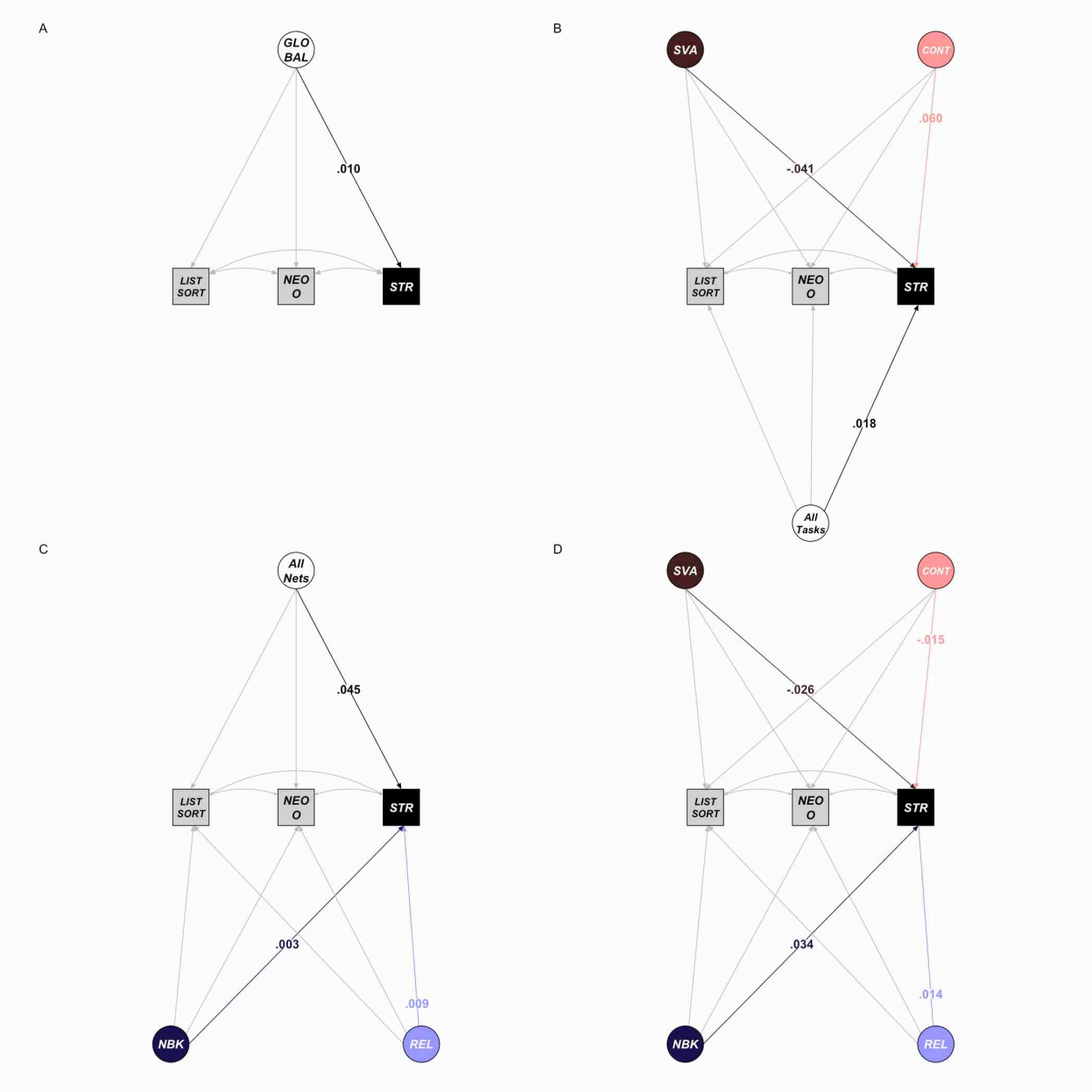
**Supplement Table 5**

| Model | List Sorting $R^2$ | PMAT $R^2$ | Strength $R^2$ |
|---|---|---|---|
| Null Model | .84% | 1.81% | .01% |
| Partial Brain Model | 6.07% | 10.75% | .58% |
| Partial Task Model | 5.25% | 9.65% | .22% |
| Full Model | 5.15% | 9.45% | .23% |

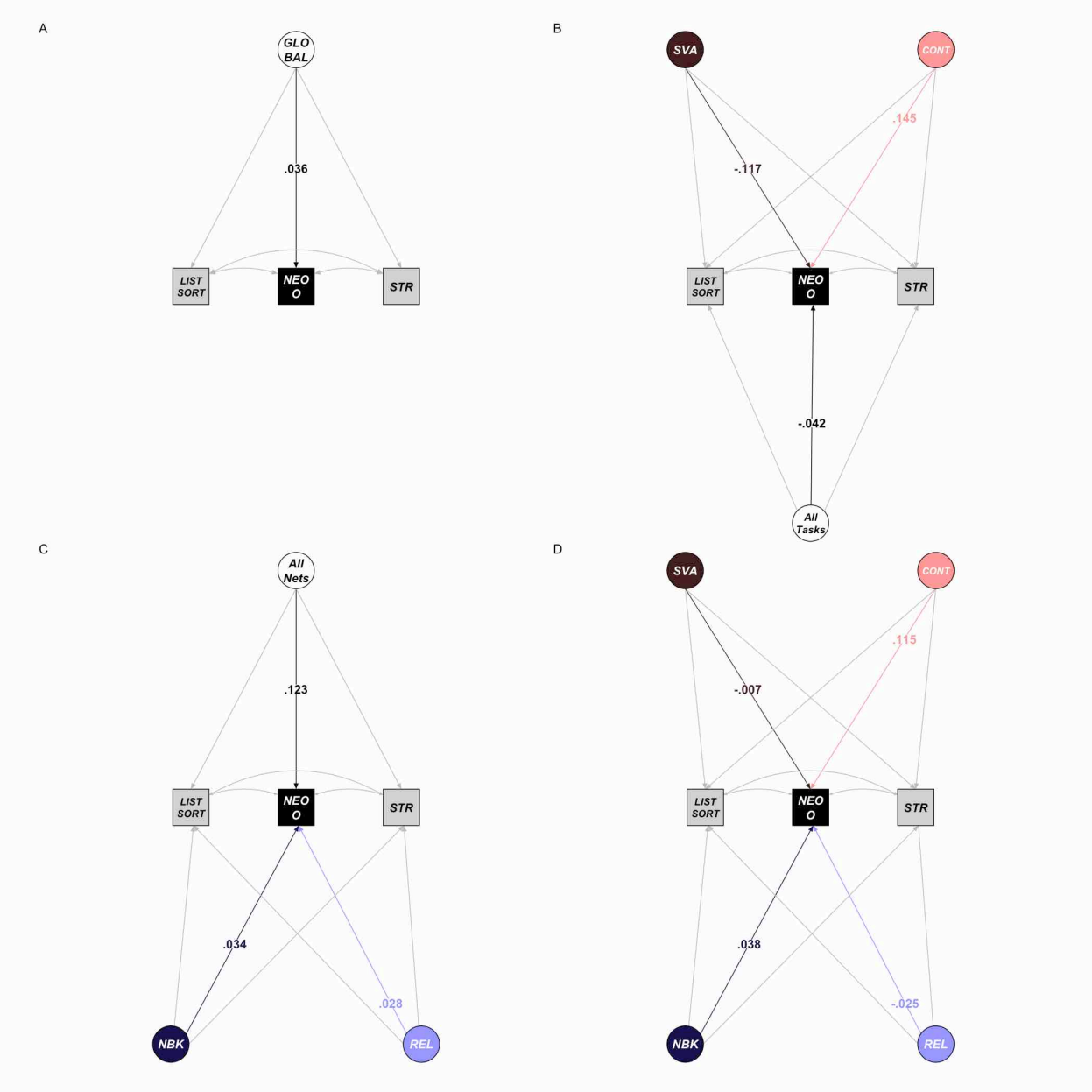
**Supplement Figure 1**



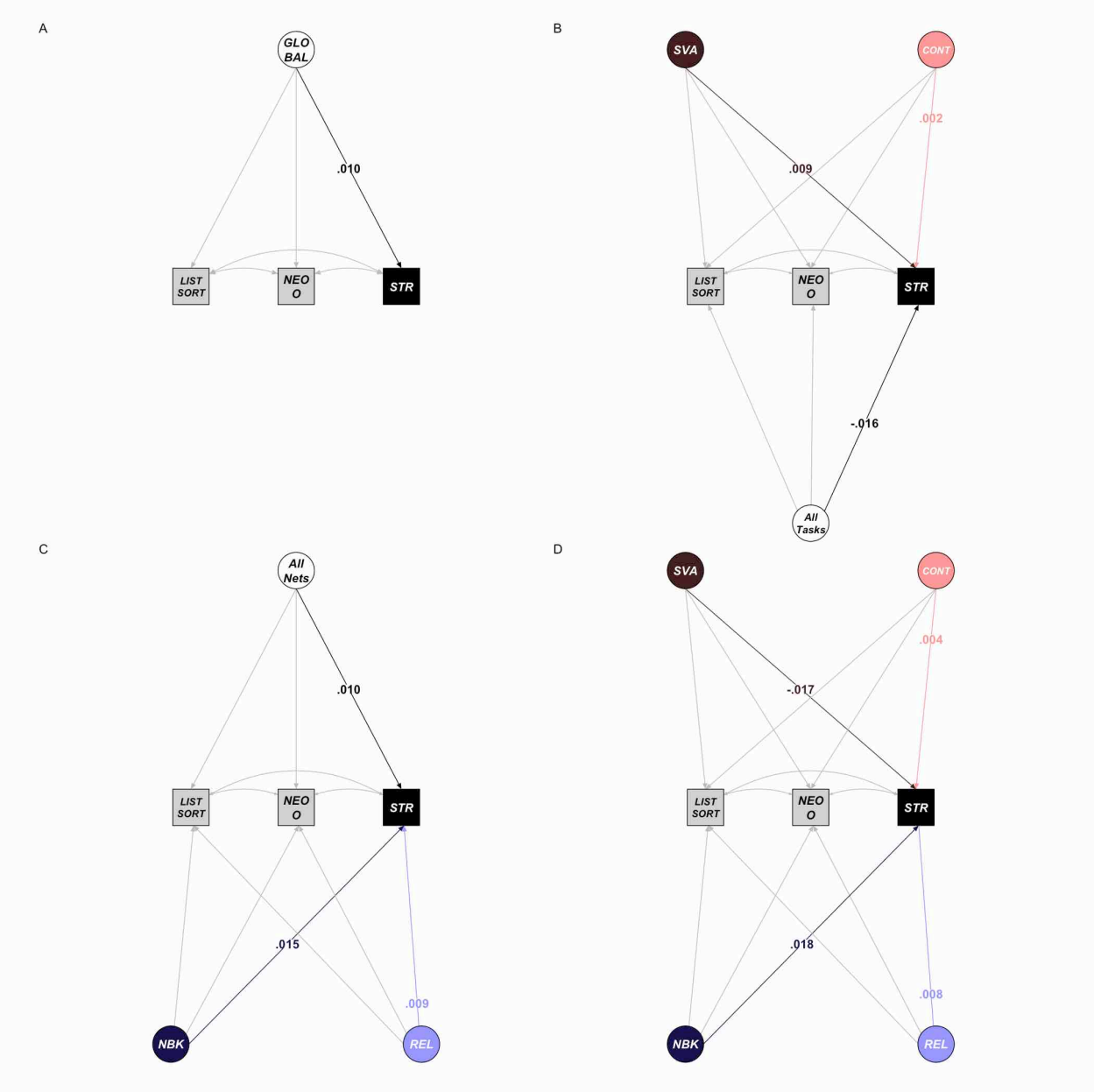
**Supplement Figure 2**



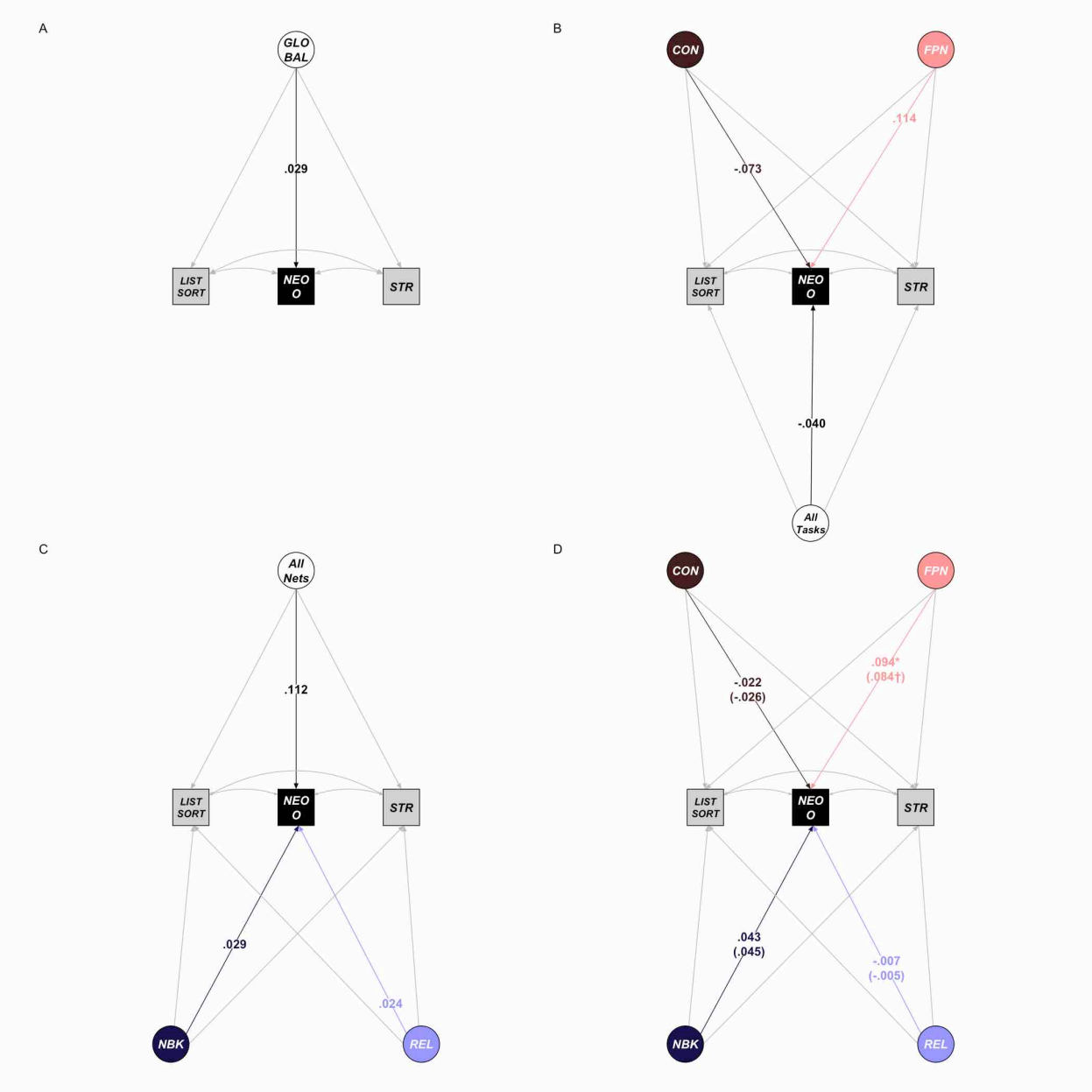
**Supplement Figure 3**



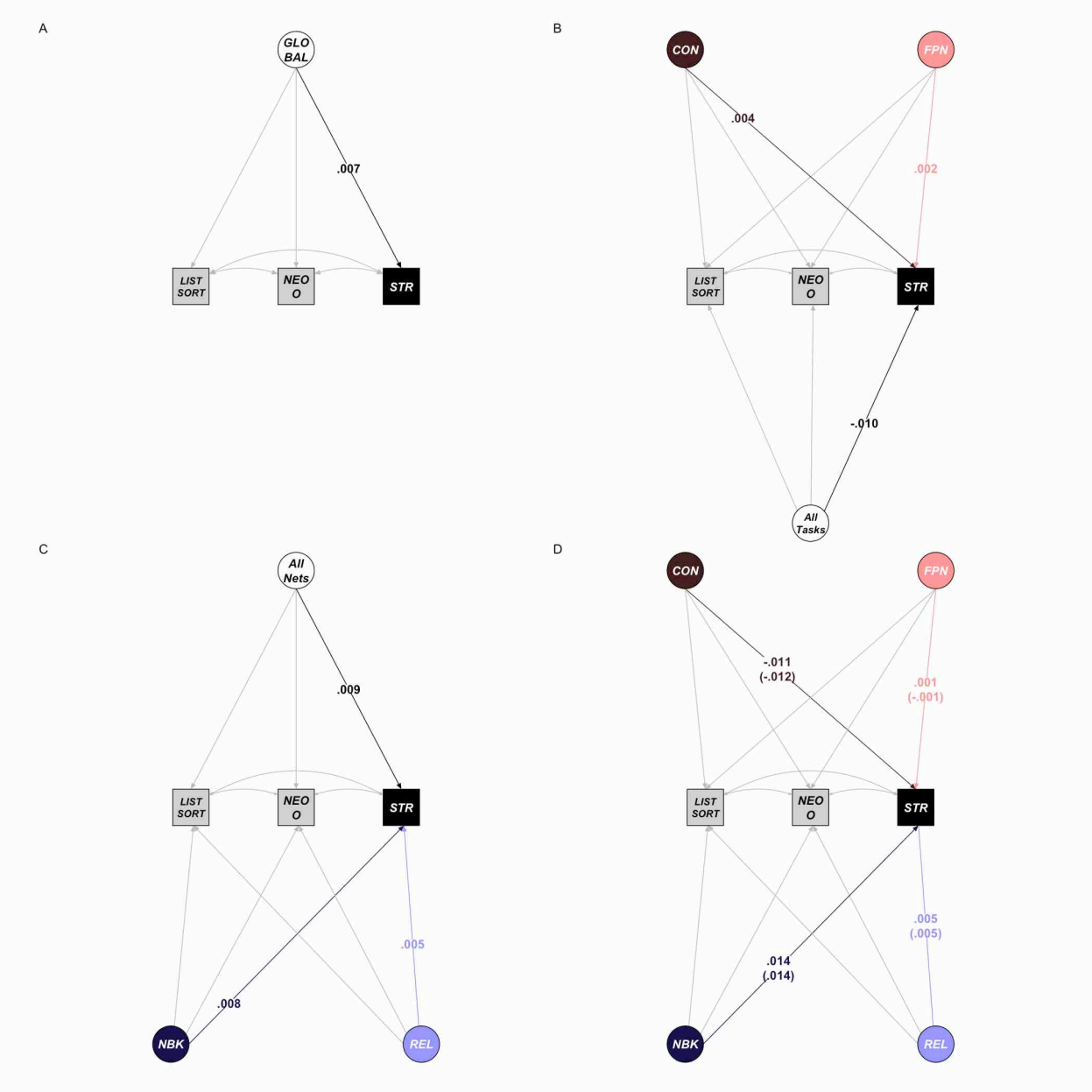
**Supplement Figure 4**



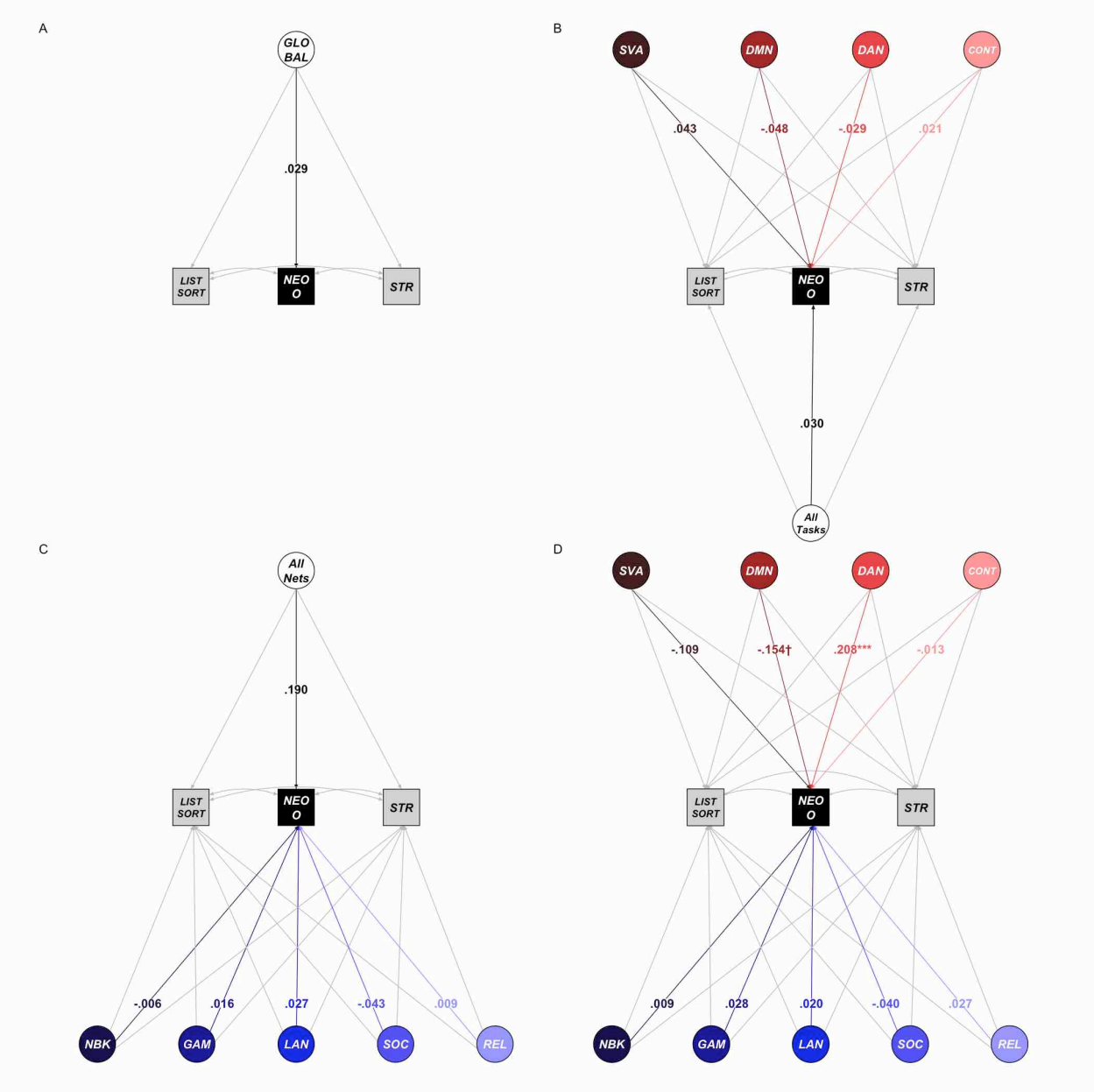
**Supplement Figure 5**



**Supplement Figure 6**



**Supplement Figure 7**



**Supplement Figure 8**